

لغويات المدونة الحاسوبية

تطبيقات تحليلية على العربية الطبيعية



د. سلطان بن ناصر المجيول
د. محمود عثمان الحاج
د. عبدالمحسن بن عبيد الثبيتي
د. عبدالله بن يحيى الفيفي
د. هند بنت مطلق العتيبي
د. عقيل بن حامد الشمري
د. محمود بن عبدالله المحمود

تحرير: د. سلطان بن ناصر المجيول

مباحث لغوية ٢٦



www.kaica.org.sa



لغويات المدونة الحاسوبية



لغويات المدونة الحاسوبية

الطبعة الأولى
١٤٣٨ هـ - ٢٠١٦م

المملكة العربية السعودية - الرياض
ص.ب ١٢٥٠٠ الرياض ١١٤٧٣
هاتف:٠٠٩٦٦١١٢٥٨٧٢٦٨ - ٠٠٩٦٦١١٢٥٨١٠٨٢
البريد الإليكتروني: nashr@kaica.org.sa

مركز الملك عبدالله بن عبدالعزيز الدولي لخدمة اللغة العربية، ١٤٣٨هـ.
فهرسة مكتبة الملك فهد الوطنية أثناء النشر
المجيول، سلطان بن ناصر
لغويات المدونة الحاسوبية./ سلطان بن ناصر المجيول
- الرياض، ١٤٣٨هـ
..ص؛ .. سم
ردمك: ٢-٩-٩٠٨٤٧-٦٠٣-٩٧٨
١ - اللغة العربية-بحوث ٢- الحواسيب-استخدام أ. العنوان
ديوي ٤١٠,٧٢ ١٤٣٨/١١٣٩
رقم الإيداع: ١٤٣٨/١١٣٩
ردمك: ٢-٩-٩٠٨٤٧-٦٠٣-٩٧٨

التصميم والإخراج
دار وجوه للنشر والتوزيع
Wojooh Publishing & Distribution House
www.wojoooh.com
المملكة العربية السعودية - الرياض
الهاتف: 4562410 الفاكس: 4561675
للتواصل والنشر:
info@wojoooh.com



مباحث لغوية ٢٦

لغويات المدونة الحاسوبية

- تطبيقات تحليلية على العربية الطبيعية -

تأليف:

د. سلطان بن ناصر المجيول
د. محمود عثمان الحاج
د. عبدالمحسن بن عبيد الثبيتي
د. عبدالله بن يحيى الفيفي
د. هند بنت مطلق العتيبي
د. عقيل بن حامد الشمري
د. محمود بن عبدالله المحمود

تحرير :
د. سلطان بن ناصر المجيول




هذه الطبعة
إهداء من المركز
ولا يسمح بنشرها ورقياً
أو تداولها تجارياً


بسم الله الرحمن الرحيم


هذه الطبعة
إهداء من المركز
ولايسمح بنشرها ورقياً
أو تداولها تجارياً

مقدمة المحرر

اعتنى هذا الكتاب بمعالجة النصوص العربية آليًا. واحتوى على ستة فصول: كل فصل قد وظّف مدونةً حاسوبية معينة بأدواتٍ منهجية لغوية حاسوبية متعددة بتعدد ديدن التحليل الآلي المنهجي في مجال (لغويات المدونة الحاسوبية corpus linguistics). وتنوعت الفصول في المناهج والأدوات بالصورة التي تجعل من كل فصل مُمَثِّلاً لمنهج ونظرية وتطبيق من عدة مناهج ونظريات وتطبيقات ضمن الحقل نفسه.

وتضمّن الكتاب ستة فصول، ففي الفصل الأول: (المعالجة الآلية للصحف العربية: تحليل الخطاب بمناهج BCU) يطرح الدكتور سلطان المجيول طبيعة المعالجة الآلية للخطاب، والمعنية بوعاء الصحف العربية في المدونة العربية التابعة لمدينة الملك عبدالعزيز للعلوم والتقنية، مع شرح المعالجة القائمة على تحليل الأنماط الخطابية. ولم يكن الغرض من الفصل تقديم تحليل شمولي لكل موضوعات تحليل تواليات الخطاب (الفوخطابي metadiscourse) بسبب محدودية الفصل، بل كان لغرض تسليط الضوء على طبيعة الاتجاهات التحليلية الخطابية القائمة على المعالجة الآلية: تحليل الخطاب بالاعتماد على مناهج بايبر وكونر وأبتون Biber, Connor and Upton، المصطلح عليها بـ BCU approaches.

وفي الفصل الثاني (المعالجة الآلية للعربية الإدارية: دراسة مدونات الإدارة وأخبار



الأسهم والاقتصاد في الوطن العربي، للدكتور محمود عثمان الحاج) جاءت تطبيقات المعالجة الآلية على نصوص الإدارة وأخبار الأسهم والاقتصاد في الوطن العربي. ويفيد هذا الفصل بمنهج تحليلي آلي يعتمد على أدوات حاسوبية مهمة لمعالجة نصوص ضخمة خاصة بالتقارير المالية للشركات في الوطن العربي)، مكونًا من ذلك ثلاث مدونات أساسية. وقد عالج مؤلف الفصل هذه المدونات بطرائق عديدة: توسيم مدونات الدراسة آليا، وأدوات معالجتها، ومنهج مقروئية النص المعتمد على الكشف عن الكلمات الطويلة والمعقدة في النصوص، وأدوات معالجة اللغة (غواص وآنت-كوك)، وسحابات الكلمات word clouds ونتائجها، وتصنيف الوثائق بواسطة تعلّم الآلة.

وفي الفصل الثالث (المعالجة الآلية للكلمات المميزة للمدونات: قضايا تقنية للدكتور عبدالمحسن بن عبيد الثبيتي) سلّط الباحث الضوء على قضايا تقنية متعلقة بحساب الكلمات المميزة keywords، وكيف أن لحساب هذه الكلمات بين المدونة الرئيسة والمدونة المرجعية أثرًا غير خاضع فقط لحجم المدونة المرجعية الذي يتطلب أن يكون أكبر بكثير من حجم المدونة الرئيسة (كما هو مقرر في أدبيات لغويات المدونة الحاسوبية)، بل برهن الباحث على وجود تأثيرات أخرى مرتبطة بالمنطقة الجغرافية، وبزمن إنتاج النص الطبيعي، وبالموضوعات.

وفي الفصل الرابع (المعالجة الآلية لمدونات المتعلمين: تحليل الأخطاء في المدونة اللغوية لمتعلمي اللغة العربي للدكتور عبدالله بن يحيى الفيفي) بيّن الباحث طرائق الاستفادة من مدونة المتعلمين التي صممها، وذلك عن طريق وضع الأسس الأولية لمفاهيم المعالجة الآلية لمدونات المتعلمين، ومناهج تحليل الأخطاء فيها، وتحديد نوعها، ووصفها، وتفسيرها، وتقويمها بشكل علمي ومنهجي.

وتقدم الدكتورة هند بنت مطلق العتيبي في الفصل الخامس (المعالجة الآلية للمدونات المتوازية واستخداماتها في تعليم اللغات وتدريب والمترجمين) مفهوم المدونات المتوازية وأهميتها في تدريب المترجمين، وتعرضُ مشروع AEPC الذي تشرف الباحثة عليه في قسم اللغة الإنجليزية بكلية اللغات والترجمة في جامعة الملك سعود، وبالتعاون مع مجموعة إيوان البحثية، مع تبيان مراحله الأولى التي استُأنفت، ومراحله الثانية التي مازالت في طور التطوير وإكمال البناء والتصميم.

ويحلل الدكتور عقيل بن حامد الشمري والدكتور محمود بن عبدالله المحمود في



الفصل السادس (المعالجة الآلية لوعاء الأخبار: تحليل الخطاب النقدي المعتمد على المدونة الحاسوبية) نصوصَ الأخبار في كل من قناتي الجزيرة والعربية حول أحداث الحرب على غزة، وقد وظفا منهج تحليل الخطاب النقدي بأدوات معالجة اللغة والتحليل الإحصائي الكمي، مفيدين حقل التحليل بمناهج معتمدة على أداة من أدوات المدونة الحاسوبية لمجال تحليل الخطاب العربي.

ولابد من التنويه إلى أن بعض الكلمات في النصوص المدونية المعالَجة في هذا الكتاب قد تضمنت أخطاء إملائية، وهي حصيلة الثقافة العربية الكتابية المعاصرة، والمليئة بالأخطاء الإملائية، وخاصةً همزتي الوصل والقطع والتاء المربوطة والهاء. وقد عُرضت هذه الأخطاء كما هي متمثلة في المدونة الحاسوبية من دون تصحيح أو تدخل؛ من أجل التمثيل الصادق للنص العربي، ومن أجل الالتزام بأدبيات لغويات المدونة الحاسوبية corpus linguistics، ومعالجة اللغة العربية Arabic language processing، واللغويات الحاسوبية computational linguistics، والإنسانيات الرقيمة digital humanities. أما تصحيح هذه الأخطاء فهو معالجة أخرى لم يُتطرق لها في الكتاب، تتعلق بتحليل الأخطاء والبحث عن دوافعها وأسبابها وكيفية معالجتها. وثمة تنويهات أخرى علميّةً، يجدر بالقارئ أخذها في الاعتبار في أثناء قراءته للكتاب، وتتعلق بالآتي:

أولاً: عدم الخلط بين المصطلحات، فمصطلح المدونة هو المدونة اللغوية وهو المدونة الحاسوبية، ومصطلح لغويات المدونة الحاسوبية هو نفسه مصطلح لسانيات المدونة اللغوية، غير أن المحرر فضّل مصطلح (المدونة الحاسوبية) بدلاً من مصطلح (المدونة اللغوية)، ومصطلح (لغويات المدونة الحاسوبية) بدلاً من مصطلح (لسانيات المدونة اللغوية)؛ نظرًا لدقتهما مفاهيميًا كما هو ديدن علم المصطلح المفاهيمي (أنظمة المفاهيم conceptual systems) الذي يرجّح العلل الدلالية والمفاهيمية دون غيرها (انظر: نقل مصطلحات اللسانيات الاجتماعية إلى العربية؛ تأليف: المحرر، مركز حمد الجاسر الثقافي، الرياض، ٢٠٠٨).

ثانيًا: عدم الاعتماد على القراءة السريعة بين الجداول وشروحاتها؛ لأن هذا العلم يحتاج إلى تأنٍ طويل في القراءة التحليلية، وتكثيف الوقت لفهم طبيعة التحليل اللغوي الآلي للغة الإنسانية مهما كانت أوعيتها، ومجالاتها، وموضوعاتها، وعلومها الاختصاصية العلمية أو الإنسانية.



ثالثًا: تطلّب هذا العلم الكثير من الجهد والوقت والتحليل العلمي من المشاركين، ووفقَ مناهج (لغويات المدونة الحاسوبية)، ولذلك فإن طلبَ الاستفادة منه يتطلب أيضا وقتًا يمتد كامتداد الجهد المبذول فيه؛ لأنه يحتاج إلى تمرّس دائم على مناهج الإحصاء المتعددة والمعالجة الحاسوبية اللغوية ضمن نظريات لغويات المدونة الحاسوبية، ومناهجها، وتطبيقاتها.

رابعًا: لكل قارئ أدركَ (أو مختصٍّ في) هذا المجال حقٌّ في طلب بيانات كل تحليل، والعمل على مبادئ التعليلية accountability أو الدحضية falsifiability أو التناتجية (التماثل والقبول) replicability العلميّة دون غيرها، والتواصل مع المشاركين عبر حساباتهم البريدية المشار إليها في (التعريف بالمشاركين)؛ من أجل الحصول على بيانات التحليل.

خامسًا: ليس الغرض من الكتاب حصر المعالجات الآلية، كونها واسعةَ المجال ومتشعبة الموضوعات بتشعّب أغراض المعالجة، إنما الغرض التعريف بطبيعة التطبيقات.

المحرر

د. سلطان بن ناصر المجيول

٨ صفر ١٤٣٨هـ

الموافق ٨ نوفمبر ٢٠١٦م

الفصل الأول

د. سلطان بن ناصر المجيول

المعالجة الآلية للصحف العربية: تحليل الأنماط الخطابية بمناهج BCU

الملخص

توظف الدراسة هنا مناهج حديثة في تحليل الخطاب discourse analysis وتحليل الخطاب النقدي critical discourse analysis بالاعتماد على ما قدمه كل من بايبر Biber وكونر Conner وأُبتون Upton والمصطلح عليه بمناهج BCU. ووظفتُ طرائق معالجات هذه المناهج الآلية على خطاب الصحف العربية، والعمل على تحليل أنماط النسوية واللانسوية أنموذجًا. ويشكل منهج BCU المرحلة الثالثة بعد مرحلتين من مراحل تحليل الخطاب؛ الأولى: مرحلة العين واليد والخطاب الصغير غير الحاسوبي، والثانية: مرحلة الاعتماد على أدوات معالجة اللغة (التكرارات والكشافات السياقية). وفي نتائج التحليل للموضوع المحدد تحليله (المرأة العربية)، شكّلت الأنماط الخطابية constructive patterns المتعلقة بالممارسات الاجتماعية-الثقافية socio-cultural practices (دون الوحدات الخطابية التي تتضمن الممارسات الاقتصادية السياسية eco-political practices) مع أو ضد المرأة-نسبيًا-صورةً متقاربة بين الإيجابية والسلبية positive and negative voices (مع رجحان أنماط الأصوات الإيجابية positive voices) من حجم مدونة الدراسة: ٢٤١ مليون ونصف المليون كلمة،



وحجم الكشافات السياقية الخاصة بالكلمة المركزية (المرأة) و(النساء) الذي يمتد إلى 31n-grams، ويتضمن ٣,٠٤٨,٤٤٧ كلمة.

الكلمات المفتاحية: المدونة الحاسوبية، تحليل الخطاب المعتمد على المدونة الحاسوبية، تحليل الخطاب بمناهج BCU، الصحف العربية، المرأة العربية.

١ . ١ . المقدمة

تنوعت اتجاهات تحليل الخطاب وتحليل الخطاب النقدي بتنوع مصادر التحليل اللغوي المنطوق أو المكتوب، فمنها ما يكون التركيز فيه على الكلمة، أو العبارة، أو الجملة التي تحقق عنصر التواصل الفعال، وذلك كله بالتركيز على المعرفة حول اللغة، وكيفية تشكلها عبر النصوص والكلام، وبالتركيز على العلاقة بين اللغة وبين السياقات الاجتماعية والثقافية المُنتجة فيها، كما يدرس الاتجاهات التي تمثل اللغة بها تصورات العالم المختلفة حول العالم وحول اختلاف الوفاقات بين منتجي اللغة وبين السياق الاجتماعي العام والخاص. ويختبر تحليل الخطاب كيفية تأثر الاستعمال اللغوي بالعلاقات بين منتجي اللغة وأثره على الهويات الاجتماعية وعلاقاتها الدقيقة، كما أنه يختبر أيضا صوريات الآراء والمفاهيم للعالم والهويات وكيفية تمثيلها في منتج لغوي اسمه «الخطاب» (بالتريدج Paltridge 2007: 2).

وتحليل الخطاب ليس حقلاً ذا علم صارم وبمتصورات تأسيسية، ومرحلية، ومن ثم كيانية؛ كونه قد بدأ قديما في سياق دراسات النصوص البلاغية والفقه لغوية إلى أن تضافرت مدارس وحقول أخرى في تكوين شعبه، وسبقه تحليل المحتوى content analysis الذي ظهر في مطلع القرن العشرين في الولايات الأمريكية المتحدة، ثم تشعبت بتشعب المدارس اللسانية والوظيفية والإثنوجرافية، بدءا من عام ١٩٥٢م عند هاريس Harris ثم نال أبعادا عديدة متضافرة التداخل التداولي والنفسي والاجتماعي والسلوكي (انظر: شارودو ومنغو ٢٠٠٨: ٤٣-٤٧)، غير أن حجم الخطاب المحلل في المجمل لا يتجاوز صفحات معدودة سواء كانت مكتوبة أو منقولة نَسْخًا من المنطوق (انظر المبحث ١ . ٢). وبعد تطور أدوات معالجة اللغات (انظر على سبيل المثال: المجيول ٢٠١٥: ٢٤٣-٢٤٥)؛ وانظر: الفيفي وأتول (Alfaifi and Atwell 2016)، ظهرت أبحاث في تحليل الخطاب تقوم على جمع نصوص رقمية تبلغ آلاف الكلمات، واعتمدت هذه الأبحاث على تحليل التكرارات والكشافات السياقية بوصفهما المنطلق لأي منهج من مناهج تحليل الخطاب (انظر المبحث ١ . ٣).



وتلا هذا الاتجاه تطوير مناهج أكثر تقدما، معتمدةً على المدونات الحاسوبية. وينطلق هذا الاتجاه من أساس مهم؛ وهو: تنوع النصوص أو المجالات (الأوعية genres أو المجالات domains للنصوص الضخمة)، وإعمال المنهج التنازلي top-down؛ أي: توزيع الأوعية، ثم إعمال المنهج التصاعدي bottom-up، ومن بعده يعمد إلى تحليل الخطاب تفصيلا لكل نص من النصوص، ومن ثمة تجزئة مجالات النصوص، ثم الموضوعات، والعودة من الأخير إلى عموم التشكل النصي للأوعية، وذلك بأدوات تحليلية تساعد على ذلك (بايبر وآخرون Biber et al. 2007، وأبتون وكوهن Upton and Cohen 2009؛ انظر المبحث ١ . ٤).

١ . ٢ . الخطاب والمدونة الورقية

بدأ تحليل الخطاب يأخذ منحاه التطبيقي المفعم أكاديميا بين الثمانينيات والتسعينيات من القرن العشرين، وقد رَسَخت أوجه تحليل الخطاب المعتمد على الخطاب الصغير أو مما لا يتجاوز عادة كتابًا واحدًا. وفي تصنيف مجلات الخطاب discourse العالمية، نجد أن أفضل مجلة علمية منها هي مجلة الخطاب والمجتمع Discourse and Society ومن بعدها مجلة الدراسات الخطابية Discourse Studies. وهاتان المجلتان لا تعدان الأبحاث في مجالات تحليل نصوص الإنجيل، والنصوص الأدبية، وتحليل أفعال الكلام، والسياسة اللغوية، والتنويعات اللغوية تحليلا خطابيا، بل كل أوراقها العلمية منصبة في مجالات العلوم الاجتماعية المتطورة، والمشكلات السياسة والاجتماعية الحديثة، والخطابات التي يندر أن تُستعمل أو أن تكون متداولة لغالب المجتمع في وسائط النقل اللغوي الحديثة (فان دايك van Dijk 1993، 1998، وفيركلوف Fairclough 1989، وفيركلوف وفوداك Fairclough and Wodak 1997). وفي نهاية الثمانيات من القرن العشرين على وجه التحديد، ظهر لتحليل الخطاب النقدي critical discourse analysis اتجاه يعتمد على تحليل استعمال اللغة والتواصل اللغوي ووحدات الخطاب فيها المكتوبة أو المنطوقة بتفسير طبيعة أثر السلطة والقوة والهيمنة الاجتماعية في تشكل الخطاب، وكيفية تواصلية هذا الأثر على تكّون خطابات جديدة. ويركز هذا المجال اهتماماته البحثية على اتجاهات محددة، من قبيل حقيقة الممارسات اللغوية الاجتماعية وطبيعة التراكيب الخطابية discursive constructions التي يتحتم تفسيرها لخطاب موقفي صغير ومحدد لا يتجاوز معظمُ بياناته صفحاتٍ معدودة، ويقبل بعض



الإحصاءات البيانية الأولية التي تعبر عن تلك البيانات الصغيرة (بور Burr 1995: 48 ،47، وانظر للمزيد مكانري وهاردي ٢٠١٦: ١٣٣-١٣٥).

وفي سياق حقل تحليل الخطاب ومجال تحليل الخطاب النقدي، فإن معظم المختصين في تحليل الخطاب يستعملون مصطلح corpus «المدونة أو المتن» بمفهومه التقليدي الورقي، وما زال أكثر محللي الخطاب التقليديين يرون المدونة بصفتها مجرد مجموعة من النصوص القصيرة والمنتجة سواء كانت منطوقة أو مكتوبة، وتمثل موقفا لغويا وسياقيًا محددا من الاستعمال اللغوي (انظر بالتريدج Paltridge 2007:156؛ قارنه بـ: كينيدي Kennedy 1998 وتوجنيني-بونيلي Tognini-Bonelli 2004 وبيكر Baker 2006 وأوربين Orpin 2005، ونزعة المدونة الحاسوبية دون المدونة الورقية في دراساتهم). ويشير بالتريدج في المصدر السابق إلى أن المدونة تكون-عادةً-متضمنة خطابًا مقروءًا حاسوبيا، ولكنه خطاب قصير يمكن التعامل معه بالعين واليد. غير أن الأبحاث المعتمدة على المدونة الحاسوبية في مجلتي discourse and society ومجلة discourse studies مازال يُتحرى فيهما في السنوات الأخيرة أن تكون البيانات أكبر من احتمال العين واليد والورق، إلا أن معظم الدراسات مازالت تتكئ على بيانات لغوية قصيرة جدًا، ويكون التعامل معها بالحدس اللغوي والتحليلي كافيًا. وثمة العديد من الدراسات والمقالات العلمية التي أُجريت ومازالت تُجرى في تحليل الخطاب وفق نصوصٍ أو حوارات قليلة الكم من جهة اللغة الخام، ووفق المنهج النوعي وبمعية المنهج الكمي النسبي المئوي لمتغيرات معدودة في النص القصير دون مناهج الإحصاء الأخرى للنصوص الضخمة.

وحال تحليل الخطاب العربي لا يعدو أن يخرج عن أمرين؛ الأول: مدونة ورقية قليلة الحجم، وممكن تحليلها بالحدس والعين واليد، أما الثاني: انصباب ركائزه على موضوعات اللغة، والحجاج، والاتساق، والانسجام، واستراتيجيات الخطاب، والتداولية (خطابي ٢٠٠٦، وقارنه مع الفصل السادس). وفي المقابل، فإن بعض المحللين للخطاب قد استرعوا الانتباه إلى تطلّعات إعمال ما يُعرف بـ «تحليل الخطاب المعتمد على المدونة (الحاسوبية) corpus-based discourse analysis، ومن ثمة إعمال ما يُعرف بـ: المنهج التنازلي top-down approach وبـ: المنهج التصاعدي bottom-up approach في الأبحاث الأجنبية في مطلع القرن الحادي والعشرين، ولم يبدأ التطلع إليهما في العالم العربي إلا مع بداية أول بحث عربي اعتمد في تحليل الخطاب النقدي على تحليل الخطاب



النقدي المعتمد على المدونة الحاسوبية corpus-based critical discourse analysis (الفصل السادس من الكتاب).

١ . ٣. الخطاب والمدونة الحاسوبية

انطلقت الأبحاث في تحليل الخطاب بالاعتماد على أدوات معالجة اللغة[١]، وعلى الرغم من بداية هذه الأدوات منذ مطلع القرن الحادي والعشرين، إلا أن قلة قليلة من الأبحاث المتعلقة بتحليل الخطاب، والخطاب النقدي، واللغة، والمجتمع، واللغة، والنفس، والتواصل قد أجريت بالاعتماد على أدوات معالجة اللغة. وانطلق بعض محللي الخطاب إلى أهمية الإجراءات الإحصائية وإعمال التحليل الآلي عن طريق مناهج أدوات معالجة اللغة، التي تعتمد أساليبها على تكشيف البيانات الرقمية المتعلقة بأسئلة ومناهج تحليل الخطاب وتحليل الخطاب النقدي (بيكر وآخرون Baker et al. 2008). ويجري ذلك عن طريق توظيف أدوات معالجة اللغة من التكرارات frequencies، والكشافات السياقية الممتدة expanded concordances، والبيانات الإضافية metadata، ومواضع المدى span positions في المكشافات السياقية الممتدة[٢] expanded concordancers وحسابات المقارنات الإحصائية بين المدونة الرئيسة والمدونة المرجعية الأولية: مربع كاي، ومعامل الغرابة، والمعلومات المتبادلة، وقياس_ت، وقياس_ز، والاحتمال اللوجارثمي، ومعامل دايس (الزهرة)، ومعامل اللوج_دايس (الزهرة اللوجارثمية)، وغيرها الكثير.

وفي تحليل بيكر وآخرين Baker et al. 2008 (انظر أيضا: بيكر ومكانري Baker and McEnery 2005) لمدونة مكون من ١٤٠ مليون كلمة من مقالات الصحف البريطانية، ناقشوا بعض المناهج الممكن دمجها في سياق تحليل الخطاب النقدي، ومدى فعاليتها ودقتها لأغراض هذا التحليل. وكان تحليلهم محصورًا على اللاجئين، وطالبي

١- لا يمكن إعادة شرح أدوات معالجة اللغة وتفصيلها. لكن: يُرجى من القارئ أن يعود إلى بحث الفيفي وأتويل Alfaifi and Atwell (٢٠١٦)، وبحث المجيول (٢٠١٥).

٢- هذه الإجراءات الممكن استردادها من أدوات معالجة اللغة مختلفة نسبيًا عن مناهج بايبر وكونر وأُبتون BCU approaches المعمول بها في المعالجة الآلية في هذا الفصل. وعلى الرغم من أن معظم أبحاث تحليل الخطاب في العالم مازالت غير مهيئة لهذا المنهج إلا أنه من الأهمية طرحه للقارئ العربي والمختصين العرب في تحليل الخطاب نظراً لاهتمامه بالتحليل الفوق خطابي والتحليل القائم على عدة أوعية ومجالات لمدونة حاسوبية آهلة لهذا العمل والتحليل.



اللجوء السياسي في بريطانيا، والمهاجرين، وطالبي الهجرة إلى بريطانيا، ويعرف هذا المشروع باسم (رايسم RASIM). ووظفوا في تحليلهم الكشف الآلي عن التصاحب اللفظي للكلمات المعنية بالمشروع بالإضافة إلى تحليل كشافاتها السياقية، وذلك كله قد كان لغرض وضع تصنيفات عامة لمعانيها الإيجابية أو السلبية، ولغرض تشجيع محللي الخطاب النقدي على استعمال أدوات معالجة اللغة لأي غرض تحليلي نقدي للخطاب، والاتجاه نحو التحليل الكمي بإزاء التحليل النوعي. ويكفي من عنوان ورقة بيكر وآخرين :

A useful methodological synergy? Combing critical discourse analysis and corpus linguistics to examine discourses of refuges and asylum seeker in the UK press

(الترجمة: التآزر المنهجي المفيد؟ دمج تحليل الخطاب النقدي ولغويات المدونة الحاسوبية من أجل تحليل خطابات اللاجئين وطالبي اللجوء في الصحف البريطانية)

دعوتها إلى توظيف مناهج تحليل الخطاب ضمن ما تفيد به أدوات معالجة اللغة، وإعمال الأخيرة في تحليل الخطاب النقدي، بدلا من انفصالهما عن بعض، الأمر الذي يجعل من تحليل الخطاب النقدي من دونه غير كافٍ أو مقنع.

وفي سياق تحليل الخطاب النقدي في البحث العربي، أجد ضرورة الإشارة إلى أول دراسة عربية من نوعها، قامت على تحليل مدونة لغوية من قناتي العربية والجزيرة حول أخبار الحرب على غزة في النصف الثاني من العام ٢٠١٤م، وهي دراسة كل من الشمري والمحمود (الفصل السادس)، وعنوانها: المعالجة الآلية لوعاء الآخبار: تحليل الخطاب النقدي المعتمد على المدونة الحاسوبية. وبلغ عدد كلمات المدونة المحللة وفقا لسياق البحث فيها ٥٢٨،٦٤١ كلمة، موزعة على ١٥٢٤ نصًا إخباريًا. وتمثل نصوص قناة العربية منها ما نسبته ٣١٪، أما نصوص قناة الجزيرة منها فتمثل ما نسبته ٦٩٪ من حجم مدونة الدراسة، واعتمد الباحثان على التحليل الكمي القائم على استخراج قوائم التكرار، والكلمات المميزة لنصوص كل قناة مع المقارنة بينهما، وتمثيل الكشافات السياقية مع تحليل الكلمات المصاحبة للكلمات الأكثر تكرار وللكلمات المميزة، وتوصلت الدراسة إلى الكشف عن المفارقات بين القناتين في نوعية الكلمات المستعملة لوصف الأحداث وصياغة الأخبار، فالعربية تميل إلى الرسمية أو العمومية على عكس



الجزيرة المهتمة بالوصف التفصيلي وانتقاء كلمات تتضمن معانٍ ثقافية وتاريخية ونقدية في سياق وصف دقيق لتفاصيل الحرب على غزة.

وفي هذه الدراسة، نلحظ أن تحليل الخطاب يمثل الاتجاه الثاني القائم على استعمال أدوات معالجة اللغة، وعلى جمع مدونة حاسوبية ممثلة لخصوصية التحليل ونوعية الخطاب، وعلى توظيف الإحصاءات التي تُظهر عدد تكرار الكلمات والمتصاحبات، وعلى الكشافات السياقية الخاصة بالكلمات المركزية الخاضعة للتحليل في المدونة. وعلى قلة الأبحاث حول هذا الاتجاه في الأبحاث العالمية، إلا أن دراسة الشمري والمحمود (الفصل السادس) تعد أول دراسة عربية قائمة على المدونة الحاسوبية الضخمة دون المدونة الورقية الصغيرة في حجمها.

وعلى مزية تحليل الخطاب والخطاب النقدي بالاعتماد على المدونة الحاسوبية وأدوات معالجة اللغة، إلا أن ثمة منهجًا أكثر اتساعا، يقوم على تحليل الخطابات المتنوعة بتنوع الأوعية والمجالات، ويقوم على تجزئة هذه النصوص، وتصنيفها وفقا لموضوعات الممارسات الاجتماعية والاقتصادية والسياسية، وموضوعات التواصل اللغوي بين الثقافات من حيث طبيعة التفاعل الناتج منها، وكل ذلك من أجل إعادة لَمْلَمة متجزئات الخطابات المتعددة من جديد، وبناء الوحدات الخطابية فيها بشكل متقابل وصولا إلى جسم المدونة الضخمة بصورة مختلفة عن جسمها الأول. وكل هذا سيُفصل في المبحث القادم وفق مناهج BCU، وبمنهج تطبيقي على الصحف السعودية متعددة المجالات والموضوعات، وفي سياق تحليل الممارسات الاجتماعية-الثقافية وتشكلها لموضوعات المرأة العربية الاجتماعية.

١ . ٤ . الخطاب ومناهج بايبر وكونر وأبتون BCU approaches

على الرغم من أن هذا المنهج في تحليل الخطاب دقيق ويتعامل مع نصوص ضخمة متعددة المجالات والموضوعات، إلا أن كثيرا من المهتمين في تحليل الخطاب قد دعوا إلى تأجيله نظرًا إلى صعوبة تطبيقه من قبل محللي الخطاب والخطاب النقدي.

واعتمد بايبر في تحليل المدونة (والخطاب) على منهج متعدد الأبعاد، multi-dimensional analysis أو ما يعرف بـ MD analysis، الذي يقوم على دراسة أنواع النصوص بالاعتماد على تحليل العامل factor analysis وكشف نسيج الخطاب المتعدد بالأوعية والمجالات. ويعتمد التحليل-ببساطة-: كشف طبيعة التنوع اللغوي



عبر مجموعة ضخمة من النصوص (بايبر Biber 1995). ويُعتمد على كشف أبعاد طبيعة النصوص بين الأوعية والمجالات والأزمنة والمناطق الجغرافية على النصوص الضخمة. وينطلق هذا المنهج-ببساطة-من أساس حاسوبي يُعالج الخطاب وتحليله بكونهما متحركيْن غير ساكنيْن. بمعنى: أن تحليل الخطاب القائم على موقف واحد لا يعني إلا موقفه الخاص، بينما تحليل الخطاب المتحرك، أو كما أسماه بايبر وكونر وأوبتون Biber, Conner and Upton 2007 (انظر للتدقيق: أوبتون وكوهن Upton and Cohen 2009) والخطابَ القائم على المنهج المساعد على دراسات الأوعية (أو الأجناس النصوصية) genres بوصفه خطابات متحركة ذات وحدات دلالية ووظيفية يمكن تحديدها من أجل تحديد أغراضها الاتصالية وعناصرها اللغوية المتداخلة.

١. ٤. ١. تطبيق تحليلي على الأنماط الخطابية لكشافات المرأة السياقية في الصحف العربية

سيتجزأ هذا المبحث إلى ثلاثة مباحث شارحة لمناهج تحليل الخطاب المعتمد على المدونة الحاسوبية وطبيعة ملف المدونة المحللة في هذا الفصل وتنقيتها باستخراج الكشافات السياقية الممتدة من الكلمة المركزية (المرأة) و(النساء) وذلك في المبحث (١. ٤. ١). أما المبحث (١. ٤. ٢.) فأجريت فيه تطبيقا سريعًا لما يمكن أن تفيدنا به أدوات معالجة اللغة عن طبيعة تكرارات الكلمتين المركزتين وطبيعة التصاحب اللفظي معها[١]. أما المبحث (١. ٤. ٣.) فسأطبق فيه مناهج بايبر وكونر وأُبتون BCU التي تنطلق إلى تجزئة المدونة إلى بنيات نصية، وتجزئة كل بنية نصية إلى وحدات خطابية، وتجزئة كل وحدة خطابية إلى أنماط خطابية ولغوية خاصة بها، ومن ثم العودة من الأخير إلى ترتيب الوحدات الخطابية داخل البنية النصية، وترتيب البنيات النصية داخل مدونة الدراسة كلها. وسيسلط التحليل على الأنماط اللغوية والخطابية المتشكلة للنسوية (مع المرأة) واللانسوية (ضد المرأة) فيما يخص كل الموضوعات الخاصة بالمرأة والمحمولة في مدونة الدراسة الضخمة.

---

١ – اعتمدت هنا على أدوات معالجة المدونة العربية Arabic Corpus Processing Tools المعروفة باسم «غواص»، النسخة ٦, ٤ (انظر: الثبيتي وآخرون Al-Thubaity et al. 2014 أ و ٢٠١٤ب).



١ . ٤ . ٢. مدونة الدراسة (الصحف العربية)

تتضمن المدونة العربية لمدينة الملك عبدالعزيز للعلوم والتقنية KACST Arabic Corpus ٨٦٧,٠٩٤,٤٢١ كلمة. وتُمثّل البيانات النصية فيها عدة أوعية ومجالات وموضوعات (انظر الثبيتي Al-Thubaity 2015). ويهمنا من هذه المدونة وعاء الصحف العربية، ومن الوعاء نتائج الكشافات السياقية للكلمة المركزية (المرأة) و(النساء)، وعلى امتداد خمس عشرة كلمة لاحقة لهما، وخمس عشرة كلمة سابقة لهما، ليكون التتابع اللفظي على صيغة 31n-grams وفي صيغة ملف مقسم بفاصلة comma delimited file. ويحتوي وعاء الصحف genre في المدونة العربية قرابة ٢٤١,٥٠٠,٠٠٠ كلمة، من عام ١٩٠٠ إلى عام ٢٠١٣م، مُشكّلةً ما نسبته ٣٤٪ من مجمل المدونة (المصدر السابق: ٧٣٧، ٧٣٨). ولم يُجر تحليل في العالم العربي للخطاب قائم على المدونة بوصفها خطابا رقميا إلا في دراسة كل من الشمري والمحمود (الفصل السادس)، وبمنهج معتمد على أدوات معالجة اللغة دون مناهج بايبر وكونر وأُبتون (BCU approaches 2007، وانظر أيضا أبتون وكوهن Upton and Cohen 2009). وعليه فإنه من الأهمية بمكان أن نعرض أقل ما يمكن أن تقدمه أدوات معالجة اللغة لهذا المدونة فيما يخص قضايا المرأة العربية في الصحف العربية في المبحث القادم، وسيلي ذلك تطبيق مناهج بايبر وكونر وأُبتون على هذه المدونة في المباحث (١ . ٦ .)، و(١ . ٧ .)، و(١ . ٨).

وبلغ عدد تكرار الكلمة المركزية (المرأة) في الصحف العربية المحوسبة في المدونة العربية التابعة لمدينة الملك عبدالعزيز للعلوم والتقنية ٦٥,٦٣٣ مرة في ٢٠,٥٩١ نصًا، وبلغ عدد تكرار الكلمة المركزية (النساء) ٣٢,٦٦٨ مرة في ١٦,٩٧٤ نصًا، من إجمالي عدد الكلمات الفعلية tokens في البالغ ٨٦٧,٠٩٤,٤٢١ كلمة، وعدد النصوص كلها البالغ ٨٥٤,٤٢١ نصًا في المدونة العربية كلها.

وقد أخذتُ في الحسبان تكرار هاتين الكلمتين بصفتهما غير مركزيّتين، كأن يتكرر ورودهما في التتابعات اللفظية للكشافات السياقية البالغة ٣٠ كلمة، إلا أن ما يهم في منطلق التحليل بأدوات معالجة اللغة العربية هو عدد تكرارها بصفتها الكلمة المركزية



أو العقدة[1] nodal item. وبعد تصفية تلك الكشافات السياقية الممتدة من مجمل مدونة الصحف العربية، بلغ عدد كلماتها ٣,٠٤٨,٤٤٧ كلمة.

١ . ٤ . ٣. تحليل مدونة الدراسة بالاعتماد على أدوات معالجة اللغة

أول خطوة في تحليل الخطاب بالاعتماد على المدونة الحاسوبية وأدوات معالجة اللغة هو الحصول على المدونة الضخمة التي تمثل حال الخطاب المراد تحليليه، أو تكوينها بجمعها بالطرق المتبعة في تكوين المدونات الشخصية في لغويات المدونة الحاسوبية، ومن ثمّ كشف قوائم التكرار الأولية سواء كانت الكلمات المركزية محددة مسبقا أم لا. ويلي قوائم التكرار النظر إلى نتائج نوعيات الكلمات، ثم توسيع التتابع اللفظي مع الكلمات وجعلها على مدى 2n-grams. كما أن طبيعة تحليل الخطاب يحدد طبيعة معالجة مدونة الدراسة، فعلى سبيل المثال: لو أن التحليل معني بالكلمات الوظيفية function words فإنه من الضروري حصرها في قائمة الشمول include list أداة معالجة اللغة. أما في سياق تحليلنا لمدونة الدراسة، وبالتركيز على الكلمة المركزية (المرأة) والكلمة المركزية النساء، فإن اختيارنا لوضع قائمة الإيقاف للكلمات الوظيفية مهمة كون الغرض من التعريف بالتحليل منطلقًا من موضوعات المرأة ومجالات الموضوعات (انظر الجدول ١ و٢). ولا يقتصر الأمر على كشف قوائم التكرارات على مدى التتابع اللفظي 2n-grams فقط؛ لأن طبيعة هذا التتابع يعد مفتاحًا لطبيعة نصوصه.

فعلى سبيل المثال: لو أنعمنا النظر في الجدول (١)، فإن أسئلة تحليل الخطاب للنسوية واللانسوية يتطلب البحث عن المتصاحب (مع: ٩١٠ مرات)، و(ضد: ٢٨٠٣)، ولو اقتصر الأمر في التحليل على المعية والضدية وحدهما لكان الاستنتاج راجحًا إلى السلبية دون أدنى شك. ولو أنعمنا النظر أكثر، فإنه من الممكن تقسيم طبيعة التصاحب اللفظي من الناحية التحليلية إلى عدة أقسام–أذكرُ بعضًا منها–من الجدولين (١) و(٢) المتضمنين أول أكثر خمسين تكرارًا للتتابعات اللفظية مع الكلمة المركزية (المرأة) و(النساء) على مدى span: 2n-grams، وهذه الأقسام هي على النحو الآتي:

١– بلغ عدد كلمة (المرأة) في مدونة الدراسة ٨٨,٥٠٤ مرة، بينما وردت كلمة (النساء) ٤٠,٣٣٠ مرة، كون تواردهما في التتابعات اللفظية للكشافات السياقية الممتدة 30n-grams مع تواردهما بصفتها الكلمة المركزية nodal item من دون متتابعات سياقية 1n-gram.



الأول: علاقة التشارك وطبيعته وكيفية تشكّل العلاقة النصية في الخطاب بين المرأة والرجل كما في أرقام التصاحبات اللفظية: (٧)، و(٤٢)، و(٤٣).

الثاني: ما للمرأة من حقوق، وطبيعة الأصوات تُجاه هذه الحقوق سواء كانت نسوية (إيجابية) كما في المتصاحب رقم (١) أو لانسوية (سلبية) كما في التصاحب اللفظي رقم (٢).

الثالث: المرأة والمشاركة المجتمعية أو السياسية، ودلالاتها في التصاحبات اللفظية ذات الأرقام: (٤)، و(٢٤)، و(٢٥)، و(٣١)، و(٣٣)، و(٣٦)، و(٤١)، و(٤٤).

الرابع: الأنماط الخطابية لطبيعة المرأة العربية في نواة المجتمعات الصغيرة (الأسرة كما في الرقمين (١٦)، و(٢٦).

الخامس: النصوص المتعلقة بالمرأة والعمل كما في رقمي التصاحب اللفظي: (١٠)، و(١٣).

السادس: النمط المعجمي للسجل اللغوي الخاص بمنطقة جغرافية محددة كما في التصاحب رقم (٣٣)، حيث إن المتصاحب (كوتة) خاص بوعاء الصحف المصرية، ومعربة من كلمة quota التي تعني النصيب السياسي المقرر قانونيا أو المقترح دستوريا للنقاش حول حدود مشاركة المرأة سياسيًا، والمنقول بالتعريب المباشر -مع وضع تاء التأنيث العربية- من تقارير دولية تخص هذا الشأن.

السابع: تحليل التصاحبات اللفظية من حيث طبيعة معانيها الناتجة (إيجابية، سلبية)، أو من حيث وضع الفاعلين في كشافاتها السياقية (هيمنة، سلطة، عادات) وفقا للموضوع theme والمحمول rhyme.

2n-grams	ت	ت_ن
١. حقوق المرأة	2591	0.09681305
٢. ضد المرأة	2298	0.08586507
٣. دور المرأة	2088	0.07801839
٤. مشاركة المرأة	1209	0.045174442



2n-grams	ت	ت_ن
٥. قضايا المرأة	916	0.03422646
٦. لجنة المرأة	822	0.030714137
٧. المرأة والرجل	783	0.029256897
٨. شؤون المرأة	710	0.026529243
٩. وضع المرأة	664	0.024810446
١٠. عمل المرأة	609	0.022755364
١١. حق المرأة	564	0.021073934
١٢. المرأة المسلمة	540	0.02017717
١٣. المرأة العاملة	493	0.01842101
١٤. صورة المرأة	483	0.018047359
١٥. المرأة العالمي	444	0.016590118
١٦. المرأة والطفل	440	0.016440658
١٧. بحقوق المرأة	415	0.015506529
١٨. تمكين المرأة	400	0.014946052
١٩. تحرير المرأة	393	0.014684496
٢٠. قضية المرأة	390	0.014572401
٢١. يوم المرأة	365	0.013638273



2n-grams	ت	ت_ن
٢٢. لحقوق المرأة	360	0.013451447
٢٣. وحقوق المرأة	342	0.0127788745
٢٤. دخول المرأة	339	0.01266678
٢٥. تولي المرأة	292	0.010910618
٢٦. المرأة الحامل	288	0.010761158
٢٧. منح المرأة	286	0.010686427
٢٨. المرأة بشكل	282	0.010536967
٢٩. حول المرأة	278	0.010387506
٣٠. واقع المرأة	277	0.010350141
٣١. تمثيل المرأة	274	0.010238046
٣٢. مكانة المرأة	265	0.009901759
٣٣. كوتة المرأة	261	0.009752299
٣٤. المرأة الريفية	259	0.009677568
٣٥. المرأة حقوقها	253	0.009453378
٣٦. وصول المرأة	252	0.009416013
٣٧. بقضايا المرأة	246	0.009191822
٣٨. حياة المرأة	238	0.008892901



2n-grams	ت	ت_ن
٣٩. لأن المرأة	238	0.008892901
٤٠. رابطة المرأة	235	0.008780805
٤١. المرأة السياسية	233	0.008706075
٤٢. المرأة والأسرة	228	0.00851925
٤٣. المرأة والاسرة	227	0.008481884
٤٤. تعيين المرأة	227	0.008481884
٤٥. بدور المرأة	225	0.008407154
٤٦. وجود المرأة	220	0.008220329
٤٧. بيوم المرأة	194	0.007248835
٤٨. المرأة أكثر	190	0.0070993747
٤٩. حصول المرأة	186	0.006949914
٥٠. جسد المرأة	185	0.006912549

الجدول (١): أول أكثر خمسين تكرارا للكلمة المركزية بامتداد 2n-grams.

وتطول مثل هذه التقسيمات بفعل قدرات تحليل المحلل التصنيفية، وبفعل قدراته أيضا في استعمال البرامج الإحصائية أو أدوات معالجة اللغة البرمجية كبرنامج آر-ستوديو R-Studio في الكشف عن تحليل دقيق لمتصاحبات لفظية محددة تخص موضوعًا بحثيا واحدًا ودقيقًا. وفي أمثلة هذه التصنيفات، نلحظ طبيعة توظيف التقسيم الموضوعي أو الجغرافي لقضايا النسوية واللانسوية (أو غيرهما بحسب طبيعة البحث وغرضه) في خطاب الصحف العربية تجاه المرأة العربية. وبتحديد نوعية هذه الموضوعات، أو

بتحديدها كلها، فإن أداة معالجة المدونة العربية المستعملة لإظهار هذه القائمة تُمكن من إظهار الكشافات السياقية الممتدة (١٥ كلمة قبل الكلمة المركزية و١٥ كلمة بعدها)، الأمر الذي يوفر مزيدًا من الكشف عن أنماط الخطاب فيها والفاعلين فيها والسياق المحمّل فيه واستراتيجات الخطاب التي تقنع من سياق وعلاقة ومقام وأساليب لغوية، وتساعد بشكل دقيق على إجراء تحليل خطاب دقيق يقوم على فرضية محددة، أو نظرية محددة من فرضيات التداوليات ونظرياتها، أو اللغويات الاجتماعية، أو اللغويات النفسية، أو الاتصال والمجتمع والثقافات.

وينطبق ما سبق آنفا على الجدول (٢) الذي يتضمن قوائم الكلمة المركزية (النساء) ومتصاحباتها على امتداد 2n-grams، والبحث عن كشافاتها السياقية على امتداد 31n-grams عند الضغط على كل متصاحب.

2n-grams	ت	ت_ن
١. النساء والرجال	711	0.026566608
٢. النساء والاطفال	709	0.026491877
٣. ضد النساء	505	0.01886939
٤. النساء والولادة	375	0.014011924
٥. النساء الحوامل	287	0.010723792
٦. النساء والتوليد	273	0.01020068
٧. أمراض النساء	260	0.009714934
٨. النساء العاملات	250	0.009341283
٩. حقوق النساء	231	0.008631345
١٠. النساء والفتيات	206	0.0076972167



2n-grams	ت	ت_ن
١١. مشاركة النساء	194	0.007248835
١٢. النساء أكثر	193	0.00721147
١٣. سجن النساء	164	0.0061278814
١٤. النساء المسلمات	160	0.005978421
١٥. جميع النساء	158	0.0059036906
١٦. أمراض النساء	153	0.005716865
١٧. قتل النساء	131	0.004894832
١٨. دور النساء	97	0.0036244176
١٩. النساء الأكثر	94	0.0035123222
٢٠. النساء المصابات	93	0.003474957
٢١. مثل النساء	88	0.0032881314
٢٢. حياة النساء	83	0.0031013058
٢٣. تمثيل النساء	82	0.0030639407
٢٤. غالبية النساء	75	0.0028023848
٢٥. فإن النساء	73	0.0027276545
٢٦. نحن النساء	72	0.0026902894
٢٧. كيد النساء	72	0.0026902894



2n-grams	ت	ت_ن
٢٨. تشغيل النساء	72	0.0026902894
٢٩. وضع النساء	68	0.002540829
٣٠. عمل النساء	67	0.0025034638
٣١. مساعدة النساء	65	0.0024287335
٣٢. النساء المعنفات	60	0.0022419079
٣٣. النساء والشيوخ	59	0.0022045427
٣٤. النساء والشباب	59	0.0022045427
٣٥. النساء الشابات	59	0.0022045427
٣٦. النساء المتزوجات	58	0.0021671776
٣٧. فئة النساء	57	0.0021298125
٣٨. اعداد النساء	57	0.0021298125
٣٩. آلاف النساء	56	0.0020924474
٤٠. النساء الجميلات	56	0.0020924474
٤١. تشكل النساء	55	0.0020550822
٤٢. النساء شقائق	54	0.002017717
٤٣. حماية النساء	54	0.002017717
٤٤. دون النساء	53	0.001980352



2n-grams	ت	ت_ن
٤٥. النساء المشاركات	53	0.001980352
٤٦. النساء يعانين	51	0.0019056216
٤٧. اغلب النساء	51	0.0019056216
٤٨. النساء يشكلن	51	0.0019056216
٤٩. قضايا النساء	50	0.0018682565
٥٠. النساء المرشحات	47	0.0017561611

الجدول (٢): أول خمسين تكرارا للكلمة المركزية (النساء) على امتداد 2n-grams.

ولا يقتصر الأمر في قوائم التصاحب اللفظي على تحديد طبيعة البحث التحليلي (لغوي، أو اجتماعي، أو سياسي، أو ثقافي، أو نفسي، أو طبي)، بل يتجاوز تحليل التصاحب اللفظي تحليلات أكثر أهمية؛ ومنها: تحليل التفضيلات الدلالية semantic preferences وتحليل النظم الدلالي semantic prosody. فلتحليل التفصيل الدلالي وجدتُ توارد التصاحب: (قضايا المرأة، الرقم: ٥، الجدول ١) مع التفضيلات الدلالية الآتية: (طفل ومجتمع وتفكير وتنمية)؛ وعليه يمكن طرح أسئلة تحليلية عن سبب هذا التفاضل المتوارد. أما لتحليل لنظم الدلالي semantic prosody فلقيتُ التصاحب: (قضايا النساء، الرقم ٤٩، الجدول ٢) مع كلمتين متناقضتين؛ الأولى: مناهضة، والثانية: معارضة. أما تحليل النظم الدلالي، فإنه يقف على المعاني النواتج مع الكلمات المتصاحبة مع الكلمة المركزية في النص، وعلى تتابعات لفظية ممتدة وطويلة متعلقة بالمعاني الإيجابية أو السلبية أو غيرهما وفقًا للمعاني الناتجة نظمًا. وبما أن كل بحث قائم على المدونة يتطلب أن يكون قائما أيضا على التحليلات الإحصائية المعمول بها في لغويات المدونة الحاسوبية، فإن هناك العديد من الأساليب التحليلية التي توفر مساحة كبيرة لطبيعة أي تحليل خطابي معتمد على المدونة. فعلى سبيل المثال: من الممكن أن نكشف



عن الكلمات المميزة الخاصة بصحف دولة عربية معينة مقارنة بصحف الدول العربية الأخرى، وذلك بحصر الكشافات السياقية لكلمة مركزية معينة أو حصر النصوص من الصحف العربية متعددة المجالات والموضوعات أو محددة المجالات والموضوعات من دولة عربية معينة، من أجل أن تكون المدونة الرئيسة أو مدونة الدراسة، وفي مقابلها نجعل نصوص صحف الدول العربية الأخرى متعددة المجالات والموضوعات أو محددة المجالات والموضوعات مدونةً مرجعية، وهذا يعين على إظهار الكلمات المميزة في مدونة صحف الدولة العربية الواحدة، التي لا تظهر أو يكون ظهورها غير كافٍ لوصفها بالمميزة (انظر الفصل الثالث من الكتاب: المعالجة الآلية للكلمات المميزة للمدونات: قضايا تقنية، وتفصيلاته الإحصائية الدقيقة، وأساليب المعالجة فيه بصورة متطورة للكلمات المميزة بين المدونة الرئيسة والمدونة المرجعية باعتبار الموضوع أو الجغرافيا أو الزمن أو الحجم).

١ . ٥. الدراسات المؤسساتية الرسميّة حول النسوية (مع المرأة/ النساء) واللانسوية (ضد المرأة/ النساء) في العالم العربي

يفيد التحليل الذي سيُجرى في المباحث القادمة على طبيعة الممارسات الاجتماعية والأيديولوجية، وأسبابها لقضايا المرأة العربية الاجتماعية والثقافية، ونظريات الأيدلوجيا ideology والهيمنة hegemony من الذكورية masculinity والتحول المجتمعي المتشكل في خطاب مدونة الدراسة منهما إلى قطبين متقاربين نوعًا ما، قطب إيجابي وقطب سلبي، وهي نتيجة جيدة إذا ما قورنت بالذكورية المهيمنة في كل مجتمعات العالم والنظريات السائدة حول ممارسة القوة من الرجل في خطابات حللها (جرامشي Gramsci 1985)، وتعود إلى العقد الثاني من القرن العشرين، وتكشف عن الهيمنة الذكورية المباشرة على كثير من مؤسسات العالم السياسية والاجتماعية والاقتصادية والمالية، وقد تقلصت هذه الهيمنة في تحليلات خطابية لنصوص متأخرة (في نهايات القرن العشرين) لتكشف عن وجود نوع من الموازنة بين عضلات الرجل المجازية وبين رقة المرأة المجازية. وعليه فإن عدم رجحان السلبية ضد النسوية والإيجابية مع النسوية في عدة خطابات مدروسة يعد-في حد ذاته- جيدًا، الأمر الذي صاغ مفاهيم واقعية من مفهوم الأنوثة الحظويّة [النسوة المتمتعات بامتياز خاص بكل سياقات حياتهن]



privileged femininity أو الأنوثة المؤكدة emphasized femininity بشكل متزن نسبيًا مع الذكورة المهيمنة hegemonic masculinity (كونل Connell 1987).

وفي سياق الإطار النظري[١] لهذا الفصل سأقف على تلك الدراسات الرسمية والخاصة بالمرأة العربية، ومنها التطبيقية والإحصائية، التي سعت إلى كشف طبيعة الممارسات الاجتماعية social practices في العالم العربي، وواقع تلك الممارسات المؤيدة لمطالب المرأة العربية في قضاياها وشؤونها الاجتماعية الخاصة (والمصطلح عليها بالنسوية في هذا الفصل)، أو المناهضة للأنثوية وغير المهتمة لقضايا المرأة العربية المجتمعية/ الثقافية بفعل الهيمنة الذكورية[٢] hegemonic masculinity (والمصطلح عليها باللانسوية في هذا الفصل).

وينوه كونيل Connell (١٩٨٧، ١٩٩٥) إلى أن تحديد هوية المرأة لا يمكن أن ينجح في ظل ارتباطه بإطار أيديولوجي مهيمن ذكوريًا. إن معظم الدراسات العربية التي أجريت في هذا المجال تميل إلى المثالية أكثر من الواقعية والتجريبية.

وأشير هنا إلى الأوراق البحثية التي وقفت عليها، والتي تعد من وجهة نظري وصفية علمية إحصائية دقيقة عن واقع المرأة القانوني والاجتماعي في عدة موضوعات حياتية: الزواج والطلاق والعمل والحقوق والمشاركة السياسية، وقد تكفل بنشرها المعهد العربي لحقوق الإنسان في تونس (مجموعة مؤلفين ١٩٩٦)، وتتضمن دراسات ميدانية في ثمانية بلدان عربية فقط، وعنوانها: «المرأة العربية: الوضع القانوني والاجتماعي»، وتضمنت أربع عشرة دراسة كتبها ثلاث عشرة باحثة وباحث واحد عن وضع المرأة العربية الأردنية، والبحرينية، والتونسية، والفلسطينية، واللبنانية، والمصرية، والمغربية،

١- لن أدخل في الإطار النظري إلى جدليات النسوية واللانسوية وتشعباتها الرأسمالية والماركسية والراديكالية، كما أنني لن أدخل إلى أدبيات متصوراتها المتعددة بتعدد المدارس النقدية والأدبية العالمية. وسيكون التركيز هنا على الدراسات المؤسساتية الرسمية؛ نظرًا لشموليتها من جهة للدول العربية؛ ونظرًا لعلاقة هذا الشمول بالتحليل القائم في هذا الفصل.

٢- ثمة مصطلح مناهض لهذا المصطلح، وهو النسوية المهيمنة hegemonic femininity الذي وضعه لأول مرة كونيل (١٩٨٧) والذي يفضل بدلا منه مصطلح الأنثوية المؤكدة emphasized femininity. ويعرّف بوردو Bordo (مجموعة مؤلفين، ١٩٩٣) مصطلح النسوية المهيمنة بسيطرة فكرة الرقّة والانسجام والرومانسية لهيئة المرأة، واستعمل المصطلح أيضا من لدن تشوي Choi (٢٠٠٠) وكرين Krane (٢٠٠١) في المجال الرياضي ليشار به إلى طبيعة رياضة المرأة التي تتمايز مع مفهوم رياضة الرجل المحددة مسبقا في ذهنية الذكورية المهيمنة hegemonic masculinity. انظر: (بيكر وإليجي Baker and Ellece ٢٠١١: ٣٩، ٥٤).



والموريتانية. ومعظم الدراسات التي أجريت في هذا التقرير الطويل تشير إلى تغلب الحظ الكبير قانونيًا واجتماعيًا بشكل نسبي لصالح المرأة، وخاصةً في الأردن والبحرين والمغرب وموريتانيا.

والتقارير العربية المؤسسية قليلة حول وضع المرأة العربية وأحوال قضاياه الخاصة بها. وفي المقابل، ثمة العديد من الدراسات والتقارير الدولية التي كُتبت بالإنجليزية، والتي تناولت المرأة العربية في سياقها الاجتماعي/ الثقافي من جهة، وفي سياقها الاقتصادي من جهة ثانية، وفي سياقها السياسي من جهة ثالثة. ويهمني منها ما ورد عن سياقها الاجتماعي والثقافي، دون الاقتصادي والسياسي لضيق تضمينها في هذه الدراسة، وتشكل الخطاب الموجه إلى موضوعاته في الصحف العربية. ففي دراسة أجريت من قبل (تومي-تابت Tohmé-Tabet 2001) استعرضت عدة دراسات سابقة حول العوائق الاجتماعية والثقافية التي تواجهها المرأة العربية، وهي عوائق ذات صلة بالموضوعات التي ستحلل في هذا الفصل بمدونة الصحف السعودية في المدونة العربية، ولم تصل في نتائجها إلى تفسير أسباب هذه العوائق، وإلى أنماط التشكل الخطابي فيها. أما في دراسة(أوفنهاور Offenhauer 2005)، فقد كانت عن المرأة المسلمة في الشرق الأوسط وشمال أفريقيا، وأجريت في هذه الدراسات تحليلات متعلقة بموضوعات عدة للمرأة حول أيديولوجيات الدور الجنسي، والنصوص الدينية حول النسوية، وحقوق المرأة، والصحة، والتعليم، والزواج، والتوظيف، والانتخابات السياسية. وانطلقت هذه التحليلات من بيانات لعدة تقارير مأخوذة من عدة منظمات دولية، هي: الأمم المتحدة، واليونسكو، والبنك الدولي، ومنظمة العمل الدولية (٢٠٠٥: ١٨ - ٢١).

وتغلبُ على هذه الدراسة الإحصاءات والأرقام، غير أن الجلي فيهما هو اعتمادها الكلي على بيانات تلك المنظمات الدولية، الأمر الذي جعلت الدراسة مستنتجةً التحول من الصور النمطية الاجتماعية الشائعة بالسلبيات عن وضع المرأة إلى النتائج العلمية الخاضعة لمعايير العلوم الاجتماعية الصارمة، وذلك باعتمادها على تكشيف حال نصف مليون امرأة عربية. ونتجت عن هذه الدراسة أن المرأة العربية قد قاربت سياقها مع سياق المجتمع العالمي المتقدم بازدياد معدلات الاهتمام بالصحة والتعليم، وبتأخير الزواج وعناء الإنجاب والحضانة والتربية، وبتقليل الفجوة بين نسب النساء العاملات والرجال العاملين في قطاع العمل كله.



والملحوظ في نتائج الدراسة التحليلية الإحصائية السابقة هو التوسط بين الإيجابي والسلبي، وعدم رجحان أحدهما على الآخر في موضوعات المرأة المدروسة فيها، وهذا يتفق مع نتائج تحليلنا في هذا الفصل حول تقارب الأصوات الإيجابية للمرأة العربية مع الأصوات السلبية لها. ولكن أوتاوي Ottaway (٢٠٠٤: ٨، ٩) يشير على أن الأدلة التجريبية empirical evidence حول أثر حضور المرأة فعليا في المجتمعات العربية تطغى عليها الإيجابيات أكثر من السلبيات، كما تطغى على بياناتها المحللة السطحية والقيل والقال بسبب رغبة الباحثين فيها إلى الانزياح إلى إيجاد النتائج الإيجابية. غير أن ما ذهب إليه أوتاوي يمكن معالجته بمنهج توسيع البيانات، واستعمال المدونات الحاسوبية الضخمة، من أجل تحليل مادة حيادية، وتزداد حياديتها بازدياد وتنوع مصادر البيانات.

١ . ٦ . أساليب مناهج الـ BCU في تحليل خطاب مدونة الدراسة

يرتكز التحليل بمناهج الـ BCU، وهو منهج بايبر Biber وكونر Connor (وأُبتون Upton 2007)؛ وانظر (أبتون وكوهن Upton and Cohen 2009) على طريقة أكثر فريدة من نوعها، ومختلفة عن سابقتها في المبحث (١ . ٤ . ٣.) إذ اعتمدت مناهجهم على الانطلاق من المدونة كلها (في هذا الفصل: المدونة كلها هي نصوص الكشافات السياقية المتمدة إلى ٣١ كلمة، وعدد الكشافات السياقية فيها ٩٨،٣٣٧ كلمة، وبحجم بلغ ٣,٠٤٨,٤٤٧ كلمة)، ثم بتجزئة مجالات الموضوعات داخل المدونة، ثم بحصر عدد نصوص كل موضوع (أو كل وحدة خطابية discourse unit؛ انظر الجدولين ٣ و٤)، ثم بالكشف عن الأنماط اللغوية linguistic patterns، والأنماط التركيبية الخطابية discursive constructional patterns، أو ما يُسمى بالتركيب الخطابي discursive construction أو الأنماط الخطابية discursive patterns، ويعرف هذا المنهج بالمنهج التنازلي top-down approach، ويليه بعد ذلك الانتقال مرة أخرى من نوعيات الأنماط اللغوية والتركيبية والخطابية في كل نص، ومن ثمة إلى النصوص وصولاً إلى المدونة كلها، ويعرف هذا بالمنهج التصاعدي bottom-up approach، وفي المبحث (١ . ٧.) شرح تفصيلي لهذه الأساليب وفق مناهج بايبر وكونر وأُبتون BCU



في تحليل الخطاب وتحليل الخطاب النقدي. ويُجرى هذا المنهج أيضا بين عدة أوعية متنوعة لكشف أنساق الأنماط الخطابية بين الأوعية المختلفة كوعاء الصحف والكتب والحوارات ووكالات الأنباء، وهو ما يعرف بتحليل الفوخطابي (الفوق خطابي أو توالي الخطابات) عبر الأوعية metadiscourse analysis across genres، ويعرف أيضا بـ interdiscourse و interdiscursive.

أما معالجة بنيات النصوص، ومن ثم معالجة وحدات الخطاب، ثم أنماط الخطاب التواصلية واللغوية الخاصة بتحليل النسوية واللانسوية فإن خطواتها مشروحة في نهاية المبحث (١. ٧). ويهمنا هنا الآن توضيح طبيعة حجم البيانيات المستردة من وعاء الصحف، وطبيعة بنيات النص من جهة ثانية، وطبيعة بنيات الموضوعات أو الأصناف المحددة مسبقا، والمتعلقة بقضايا المرأة الاجتماعية التي يراد منها تحليل الممارسات الاجتماعية الإيجابية معها (النسوية) أو السلبية ضدها (اللانسوية). ويوضح الجدول (٤) أنواع بنيات النص text structures وحجمها من حيث حجم الموضوعات المتعلقة بالمرأة والنساء، كما يوضح أصناف وحدات الخطاب مع حجم نصوصها وحجم كشافاتها السياقية وحجم الكشافات المعنية بالتحليل.

حجم المدونة العربية	حجم الوعاء	تكرار_المرأة	عدد النصوص	تكرار_النساء	عدد النصوص
٨٥٤,٤٢١ نصًا	٢٩٥,٠٠٠ نصًا ٢٤١,٥٠٠ كلمة فعلية	حجم مدونة الدراسة (تحليل النسوية واللانسوية)			
٨٦٤,٠٤٩,٥٤٣ كلمة فعلية		٦٥,٦٥١ كلمة	٢٠,٥٩١ نصًا	٣٢,٦٨٦ كلمة	١٦,٩٧٤ نصًا
٧,٩٩٦,٦١٧ كلمة نوعية		مجموع الكلمات: ٩٨,٣٣٦		مجموع النصوص: ٣٧,٥٦٥	

الجدول (٣): حجم بيانات الكشافات السياقية للمرأة والنساء والوعاء (الصحف العربية) والمدونة العربية Arabic Corpus.

بنية النص	حجم الموضوعات (بعدد الكشافات)	مجموع الكشافات السياقية	وحدات الخطاب	مجموع النصوص	مجموع الكشافات السياقية الكلية	مجموع الكشافات المعنية
اجتماعي/ ثقافي	١٠,٩٧١		أخبار			
اقتصادي	٣٦٣	١٢,١٥٤	مقالات	٣٧٥٦٥	٩٨,٣٣٧	١٠,٩٧١
سياسي	٨٢٠		تقارير			

الجدول (٤): تجزئة ضخامة بنى النصوص text structures ووحدات الخطاب discourse units فيها لغرض تحليل النسوية (مع المرأة العربية) واللانسوية (ضد المرأة العربية).

وحري القول إن هذه الدراسة-كما أشرنا إلى ذلك في الملخص والمقدمة-ستركز على أصناف الممارسات الاجتماعية النسوية واللانسوية في سياق الموضوعات الاجتماعية الخاصة بالمرأة العربية، والمحددة مسبقا (الزواج، والطلاق، والخلع، والحضانة، والعنف، والتعليم، والتمييز، والحجاب، والرياضة، والسفر بدون محرم، وقيادة السيارة، والتحرش، ونزع ولاية الرجل؛ انظر المبحث ١ . ٧.) كونها الأكثر ضخامة من حيث الحجم مقارنة بالنصوص التي تتضمن الإطار الاقتصادي (كالتوظيف والميراث والنفقة) أو السياسي (كالانتخابات الوزارية والبلدية والمشاركة السياسية) وتشكل الخطاب النسوي واللانسوي فيها[١].

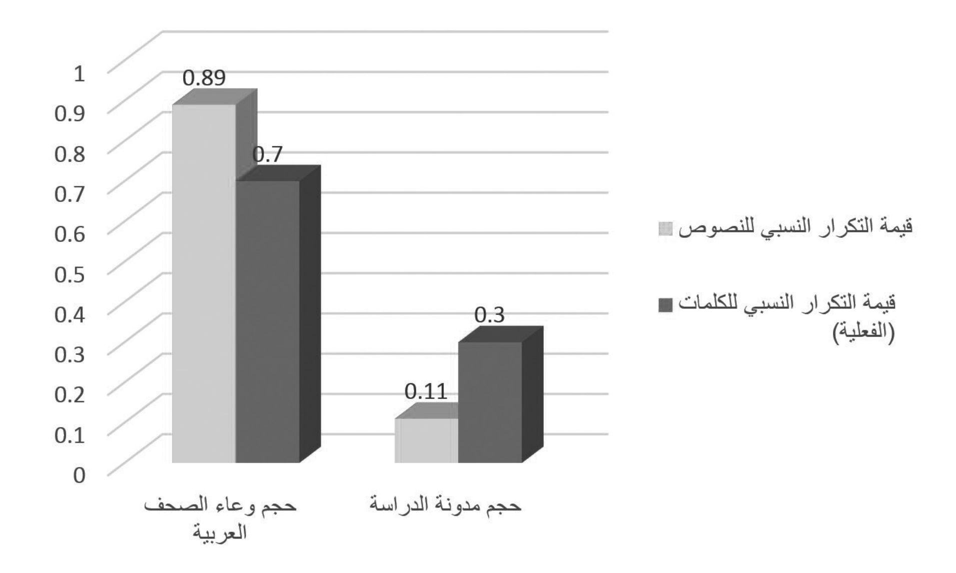
١- المفاجأة هنا هي أن موضوعات توظيف المرأة، والميراث، والنفقة بعد الطلاق للجوانب الاقتصادية، وموضوعات الترشيح الوزاري والبلدي والمشاركة السياسية للجوانب السياسية لم تنل حظًا ضخما من الظهور في وعاء الصحف العربية في المدونة العربية KACSTAC. وهو أمر غير ممكن لوعاء الصحف العربية، غير أن المشكلة قد تبدو نابعة من طبيعة المدونة الراصدة monitor corpus التي تجمع النصوص الضخمة بشكل عشوائي، دون الاعتماد على مدونة سانحة opportunistic corpus يتصدى فيها المحلل إلى نصوصِ موضوعٍ (نزع ولاية الرجل أو الأب) مثلا وبشكل محدد، أو إلى نصوصِ مجالٍ (الاقتصاد) بشكل دقيق؛ من أجل أن يُوفر أكبر عدد ممكن من التكرارات (مكانري وهاردي ٢٠١٦: ١١، ٢٠). ولعلي هنا أشير إلى ملحوظة مهمة حول موضوع محدد فيها ولافت للانتباه. عدد الكشافات السياقية المتعلقة بموضوع الميراث-على سبيل المثال-قد بلغت ٩٣ (٩٣×٣١ = ٢٨٨٣ كلمة) وأن ١٩ كشافا سياقيا قد تضمن حالتين غريبتين حول تقسيم الميراث الثابت وفق النصوص القرآنية، وكما هي مقررة في علم الميراث أو الفرائض. وهاتان الحالتان هما أن تقسيم الميراث في جمهورية مصر العربية مختلف وغير ثابت بين المدن والقرى. أما الحالة الثانية: فهي أن المرأة في فلسطين لا تستحق الورث على الإطلاق، وهو عرف قد جرت العادة عليه في المجتمع الفلسطيني، ولن أدخل هنا في البحث عن أسباب هاتين الحالتين لعدم ارتباطهما بمدونة والفصل وتحليلاته.



واعتمدتُ في هذا الفصل على أساليب آلية أعانت على تجزئة موضوعات المرأة وطريقة تشكلها في الخطاب المتمثل في الكشافات السياقة الممتدة (٣١ كلمة: 31n-grams)، وبلغ عدد هذه الكشافات ٩٨,٣٣٧، وحاصل عدد الكلمات فيها بضربها في ٣١ كلمة يساوي ٣,٠٤٨,٤٤٧ كلمة تقريبا. وفي الشكل (١) قياس نسبة حجم النصوص للكلمتين المركزيتين (المرأة) و(النساء) إلى نسبة حجم نصوص وعاء الصحف العربية في المدونة العربية (٠,١١ أي: ١١٪ و٠,٨٩ أي: ٨٩٪ على التتابع)، وقياس نسبة حجم الكلمات الفعلية للكشافات السياقية الممتدة للكلمتين المركزيتين إلى نسبة حجم الكلمات الفعلية في وعاء الصحف العربية في المدونة العربية (٠,٧٠ أي: ٧٠٪ و٠,٣٠ أي: ٣٠٪ على التتابع).

وتحليل هذا الكم من الكشافات السياقية بالاعتماد على العين واليد مستحيل، أما بالاعتماد على الاتجاه الثاني، وبأدوات معالجة اللغة التي تعين على معالجة النص بإظهار التكرارات والمتصاحبات والكشافات السياقية فإن العادة قد جرت على تحديد سؤال بحثي محدد، أو وصف النتائج الكلية من حيث التكرارات والكلمات المميزة، ومن ثمة إظهار الكشافات السياقية الخاصة بذلك الموضوع، وفي حال كون الكشافات السياقية طويلة، فإنه يُعمد عادة إلى اختيار أول مائة كشاف سياقي مثلا، أو إظهار نتائج التتابعات اللفظية على مدى خمسة متتابعات 5n-grams وذلك بتخصيص كلمة الموضوع البحثي، والعمل على وضع قائمة شمول من أجل إدراج كلمتي (المرأة) أو (النساء) فيها، ليكون نتائج البحث من هذا الكم الهائل من المدونة الحاسوبية محددًا جدًا لسؤال البحث الفرضي.





الشكل (١): حجم بيانات الكشافات السياقية للمرأة والنساء مقارنة بحجم الوعاء (الصحف العربية) ثم بحجم المدونة العربية Arabic Corpus.

١ . ٧. التحليل بمناهج الـ BCU التنازلية والتصاعدية

ينطلق التحليل بالاعتماد على المنهج التنازلي من بنيات النصوص، ففي كل بنية نصية وحدات خطابية، وفي كل وحدة خطابية أنماط لغوية وخطابية، ويُسمى هذا المنهج بالمنهج التنازلي top-down approach. ويعود هذا المنهج بالاعتماد على المنهج التصاعدي bottom-up approach بما يكشفه من أنماط لغوية وخطابية إلى وحدات خطابية معاد ترتيبها وفق طبيعة الأنماط، ثم يعود بالوحدات الخطابية إلى البنيات النصية المعاد بلورتها وفقًا لبنية النص، وصولا إلى تكوين مدونة مرتبة وفق تعدد الأنماط الخطابية وعبر مجالات domains وموضوعات topics الصحف العربية[١]. وأول ما يُتطلب إجراؤه وفق مناهج بايبر وكونر وأبتون هو تحديد مجموعة من الأصناف الخطابية الخاصة بالمرأة (كالموضوعات في سياق هذا الفصل).

١ – أطلق فان دايك على هذه الطريقة في نهاية الثمانيات مصطلح schemata ومصطلح superstructures نظرا لتركيزه على النمط التنظيمي الكلي أصناف الخطاب المتنوعة (فان دايك van Dijk ١٩٨٨ : ٢٦). وقد كانت فكرةً يصعب تحقيقها في ظل عدم إمكانية توظيف أدواتها التقنية والبرمجية والآلية التي تعين على تطبيقها بشكل دقيق.



وعدد الكشافات السياقية الخاصة بالموضوعات الاجتماعية الثقافية المحددة مسبقا (الجدول ٥) ينقص عن عدد الكشافات السياقية في الجدول (٤)، والبالغ عددها: ١٠,٩٧١، وبفارق ٨٤ كشافًا سياقيًا غير ذات صلة بالموضوعات، كرصد كلمة (النقاب) وإقصائها لارتباط ياء النسبة وتاء التأنيث فيها (النقاب+يّة). والسياق الخاص بموضوعات المرأة أو النساء في الإطار الاجتماعي/ الثقافي شكلت حجمًا كبيرًا، وبواقع ١٠,٨٨٧*31n-gram؛ أي: ٣٣٧,٥٥٩ كلمة (انظر الجدول ٥).

ومدونة الدراسة هنا قد عولجت بمنهجي التنازلي top-down والتصاعد -bottom up، فالانطلاق من المنهج التنازلي ساعد على كشف طبيعة وحدات الخطاب في مدونة الدراسة، واختيار كل النتائج المتعلقة بموضوع تحليل الخطاب (المرأة والنساء/ النسوية واللانسوية في خطاب الصحف العربية)، وتوظيف المنهج الثاني حدد الموضوعات التي ستحلل والمتعلقة بالكلمة المركزية (المرأة) و(النساء) وبامتداد تتابعتها (الخطابية) من ٩٨,٣٣٧ كشافا سياقيا، وبتقليصها وفقا للموضوعات المحددة مسبقا (الجدول ٥) ليكون عدد كشافاتها السياقية (١٠,٨٨٧) كشافا.

بنية النص	الموضوعات ذات الصلة بالتحليل		عدد الكشافات المستردة 31n-grams
اجتماعي/ ثقافي	الحالة الزواجيّة (٣٢٦٥ كشافًا سياقيًا)	الزواج	٢٤٠٦
		الطلاق	٥٨١
		الخلع	١٤٣
		الحضانة	١٣٥
	العنف الأسري		٢٨٧٠
	التعليم		١٦٧٤
	التمييز ضد المرأة		١٣٤٨
	الحجاب/ النقاب		٨٧٤



بنية النص	الموضوعات ذات الصلة بالتحليل	عدد الكشافات المستردة 31n-grams
اجتماعي/ ثقافي	الرياضة النسائية	٤٥٩
	السفر من دون محرم	٢٦٦
	الوصي (الرجل)	٨٦
	قيادة السيارة	٨٤
	التحرش الجنسي	٤٢
	نزع ولاية الرجل	٣
مجموع عدد الكشافات السياقية: ١٠,٨٨٧، ومجموع عدد الكلمات: ٣٣٧,٥٥٩ كلمة (مدونة الدراسة وفق المنهج التنازلي top-down approach)		

الجدول (٥): عدد الكشافات السياقية للموضوعات المحددة من البنية النصية (اجتماعي/ ثقافي) عبر الوحدات الخطابية الثلاث: الأخبار، والمقالات، والتقارير، وحجم مدونة الدراسة النهائية.

أما خطوات مناهج الـ BCU بالمنهجين التنازلي والتصاعدي في تحليل اتجاه سيرورة الخطاب (أو تواصلية الخطاب) في الخطاب المدوني كبير الحجم منذ أول خطورة حتى آخر خطوة فنوجزها على النحو الآتي:

الخطوة الأولى: تنقية حجم النصوص الكلية بواسطة استخراج الكشافات السياقية المحصورة كلماتُها المركزية بكلمتي (المرأة) و(النساء) من وعاء الصحف من المدونة العربية.

الخطوة الثانية: حفظ الكشافات السياقية الممتدة (٣١ كلمة 31n-grams) في ملف محفوظ بنوع type: المقسم بفاصلة comma delimited.

الخطوة الثالثة: فرز الكشافات السياقية من وحداتها الخطابية: (الأخبار الصحفية، والمقالات الصحفية، والتقارير الصحفية) وفقا لبنياتها النصية: (اجتماعي/ ثقافي، واقتصادي، وسياسي) بالكلمات: اجتماع/ ثقافة، واقتصاد، وسياسة.



الخطوة الرابعة: تكشيف الملف السابق بواسطة الكلمات المفتاحية، أي: الموضوعات المتعلقة بالسياق الاجتماعي الثقافي للمرأة: الزواج، والطلاق، والخلع، والحضانة، والعنف، والتعليم، والتمييز، والحجاب، والرياضة، والسفر بدون محرم، وقيادة السيارة، والتحرش، ونزع ولاية الرجل)، وإبعاد ما ليس مرتبًطا بالموضوع، فكلمة (العنف) مثلا قد وردت في مدونة الدراسة ٣١٤٦ مرة، بينما المستخلص منها لغرض التحليل في هذا الفصل قد بلغ ٢٨٧٠ كونه خاليًا من المتصاحبات (الأسري) أو كون الكلمة المركزية (المرأة) أو (النساء) قد ورد في كشافهما السياقي الممتد 31n-grams المتصاحب (العنف) لسياق لا يتعلق بالمرأة أو النساء بحد ذاتهما.

الخطوة الخامسة: تكميم عدد الكشافات السياقية المعنية بالموضوعات فقط باستخراجها بواسطة كلمات الموضوعات المراد تحليلها خطابيًا، والمذكورة في الخطوة الرابعة.

الخطوة السادسة: فرز الكشافات السياقية للموضوعات المراد تحليلها خطابيًا، وكل موضوع في ملف مستقل مفرز بواسطة برنامج Notepad++ (انظر الشكل ٢).
وفي تحليلي في هذا الفصل، استبعدتُ الملف المتعلق بالنسوية واللانسوية في الاقتصاد والسياسة نظرا لقلة حجمهما مقارنة بحجم الموضوعات الاجتماعية/ الثقافية في الصحف العربية في المدونة العربية.

الخطوة السابعة: فرز الكشافات السياقية المتعلقة بملف البنية النصية الاجتماعية الثقافية للنسوية واللانسوية بحسب الموضوعات الأربعة عشر، كل موضوع يفصل بكود Extended (/n, /r, /t, /0, /x...) بعد تعليمه بكلمات الموضوع نفسه (الزواج، الطلاق، الخلع، إلخ. انظر: الشكل ٣).

الخطوة الثامنة: تكميم (تحديد الكمية أو الحجمية sizeability) الأنماط التركيبية الخطابية للنسوية واللانسوية داخل كل ملف من ملفات الموضوعات الأربعة عشرة.

الخطوة التاسعة: عرض الأنماط الخطابية للنسوية واللانسوية، وتحليلها وفقًا لكمية حجم النسوية وكمية حجم اللانسوية وكمية حجم ما ليس متعلقا بالنسوية واللانسوية، وذلك لغرض تحليل أنماط الخطابية وفقا للمذكور.

الخطوة العاشرة: إعادة فرز الأنماط الخطابية بين ما هو مع المرأة وبين ما هو ضد المرأة فيما يتعلق بموضوعاتها الاجتماعية والثقافية وصولا إلى كشافات سياقية معاد تنظيمها وفقا لذلك.

woma(e)n-15+15.csv

Search "الطلاق" (581 hits in 1 file)

F:\Post-doctorate\إنجليزي\Women in KACST Arabic corpus\Data\Files\woma(e)n-15+15.csv (581 hits)

Normal text file   length : 16314961   lines : 98338   Ln : 1   Col : 36   Sel : 0 | 0   Dos\Windows   ANSI   INS

الشكل (٢): مثال لاستخراج كشافات الكلمة المركزية (المرأة) و(النساء) لموضوع (الطلاق) في برنامج Notepad++.

الطلاق_المرأة_النساء.txt - Notepad

File Edit Format View Help

Search "الطلاق" (581 hits in 1 file)

F:\Post-doctorate\إنجليزي\Women in KACST Arabic corpus\Data\Files\woma(e)n-15+15.csv (581 hits)

Line 108: / Line 523: / Line 529: / Line 1158: / Line 1160: / Line 1212: / Line 1212: / Line 1215: / Line 1241: / Line 1281: / Line 1477: / Line 1579: / Line 2681: / Line 3207: / Line 3229: / Line 3229: / Line 3230: / Line 3230: / Line 3325: / Line 3550: / Line 3701: / Line 4421: / Line 5607: / Line 5660: / Line 6045: / Line 6289: / Line 7129: / Line 7148: / Line 7149: / Line 7252: / Line 7555: / Line 7713:

الشكل (٣): تصفية أسطر الكشافات الخاصة بالكلمة المركزية لموضوع (الطلاق) في مدونة الدراسة.

وتتطلب هذه المناهج الوقت والجهد لإتمامها، بدءًا بتجزئة الكشافات السياقية



للمرأة والنساء في الصحف العربية إلى وحدات خطابية محددة، ومن ثم الكشف عن أنماط تواصلية أو لغوية أو خطابية محددة (المنهج التنازلي: من الخطوة الأولى إلى الخطوة التاسعة) وانتهاءً بجمع الأنماط إلى بنية وحدات الخطاب بشكل مختلف في الجمع والتصنيف عن الأول (المنهج التصاعدي: الخطوة العاشرة). وثمة اختلاف طفيف بين منهج (بايبر وكونر وأبتون Biber, Connor and Upton 2007) وبين المنهج الذي أجريته هنا، وهو أن تواصلية الخطاب بين النصوص ووحدات الخطاب تتضمن الاهتمام بالتصنيفات الوظيفية والتواصلية communicative/functional categories بينها، وبالتحليل اللغوي تبعا لخصائصه النحوية والمعجمية. أما ما أجريته في التحليل فهو متعلق فقط بالنسوية واللانسوية (المبحث ١ . ٨).

١ . ٨. نتائج تحليل النسوية واللانسوية لموضوعات المرأة في مدونة الدراسة

جُمعت الأنماط الخطابية للنسوية واللانسوية وفقا لكل موضوع مع الإشارة إلى حالات الأنماط الخطابية دون سياقات الفاعلين فيها وخلفياتهم الثقافية ونوعياتهم ما يتمتعون به من سلطة اجتماعية. والسبب في عدم الخوض في نوعيات الفاعلين يعود إلى ما لتخفيتها من أهمية في كشف الحالات دون الفاعلين، والوقوف على الحالات دون التأثر أو التأثير بنوعيات الفاعلين. كما أن بعض الموضوعات قليلة التكشيف في مدونة الدراسة من قبيل قيادة السيارة، والوصي، والتحرش، ونزع ولاية الرجل متعلقة بالمجتمع السعودي دون المجتمعات العربية الأخرى.

ففي الموضوعات العامة في الجداول (٦)، و(٧)، و(٨)، والمتعلقة بالمرأة نورد طبيعة الخطاب بتلك الموضوعات وما تحمله من حالات إيجابية (الجدول ٦)، أو سلبية (الجدول ٧)، أو غير ذات صلة بمحمل الإيجابية أو السلبية، مع ذكر عدد تكرار كل حالة لكل موضوع. ويوضح الجدول (٦) عدد مرات الحالات الإيجابية (الأصوات الإيجابية للنسوية) للموضوعات المحللة، وسنحلل محامل rhymes الموضوعات themes التي تكشف طبيعة التشكل الفوخطابي. ففي الأحول الزواجيّة: (الزواج، والطلاق، والخلع، والحضانة) تشكّلُ خطابًا إيجابيًا وسلبيًا معًا تجاه نسويتها، يكمن في أحقية المرأة بإدارة زمام شؤونها الخاصة دون أية استثناءات. وفي أنماط الخطاب المتعلق بالخلع، فإن الأنماط الإيجابية فيها تُشكل مطلق الحرية دون أية قيود للنسوية، وقد رصدتُ بعض الفتاوى



التي تؤكد ذلك بشرط تفعيلها من منطلق الفتوى دون منطلقات دولية كاتفاقيات سيداو من الأمم المتحدة. أما أنماط (الحجاب) و(النقاب) فهي ليست ملزمة في عدد تكرارات حالتها في مدونة الدراسة، كما أن ثمة أصوات تدعو إلى تزيين صورة الحجاب بالألوان والتطريز، غير أن العجيب في بعض الأنماط هو ظهور تصور مجتمعي عربي بأن الحجاب المطرز مدعاة إلى فعل التغزل أكثر من كونها غير محجبة.

وفي عدد تكرارات حالة أنماط ممارسة الرياضة، فإن العدد المذكور قد حمل نمط اللاعلمية في دعوى من يقول بإحداثها أضرارا جسدية فسيولوجية/ بيولوجية للمرأة. وفي أحوال «المحرم» وإلزاميته من فاعلين معروفي الهوية للمحلل، فإن عدد ما تكرر قد حمل المناهضة ضد ذلك، مع ضرورة جعله خيارا للمرأة التي ستطلب وجوده غالبا في حال عدم شعورها بالأمن عند الانتقال.

وفي عدد التكرارات الخاصة باللانسوية أو الأصوات السلبية حول الموضوعات في الجدول (٧)، فإن معظم تشكلها نابع من السلطة الدينية المتغايرة في مواقفها ومن الهيمنة الذكورية المجتمعية، فعلى سبيل المثال: نجد تعدد الزوجات خيارًا مفتوحًا للرجل، كما أن الطلاق ليس خيارًا للمرأة، وثمة حالات من زواج القاصرات متروك أحوالها لخاصتها من العائلتين. وفي موضوع الحضانة، تم تكميم حالات عدم أحقية المرأة في تربية الأولاد، وفي حال العنف، تمثلت في مدونة الدراسة أحداث من العنف المقبولة من وجهة نظر الشريعة الإسلامية، ويختلف هذا في الدراسات الاجتماعية للنسوية مع ما أسماه (بورديو Bourdeiu 1991) بالقوة الناعمة soft power التي تخيب أمل الهيمنة بين الجنسين، والتي تأخذ أشكال الإغراء والإقناع بالمال والجاذبية المادية. أما فيما يتعلق بموضوع التعليم، فقد رصدتُ ٦٣ صوتًا ضد تعليم المرأة التي تضطر في تعليمها إلى السفر بعيدًا عبر وسائل النقل العامة. أما الأصوات السلبية فيما يخص عدد تكراراتها للخلع، فقد تشكلت من حجة إهانة «الرجولة» كون مصير الخلع بيد المرأة يشجع من حالات الانفصال لأسباب اعتيادية، كما أن هذه الأصوات قد شجبت كثيرا احتفال النساء ممن خلعن أزواجهن بشكل رسمي وعلني في المجتمعات العربية. وفي موضوع التمييز، ثمة ثلاثة كشافات سياقية ممتدة تحمل موضوع عدم إمكانية تلاشي الهيمنة الذكورية في المجتمعات العربية كونها في الأصل مازالت قوية في المجتمعات الغربية والشرقية بأشكال مختلفة، و٣٤٧ كشاف سياقي حمل حالات عدم قدرة المرأة



على دفع التمييز عنها في المحاكم، وعدم تفعيل اتفاقيات سيداو التي أُقرّت منذ ثلاثين عامًا. أما بقية الكشافات السياقية فهي تشير إلى أن اتفاقية سيداو ليست جديدة على القانون الشرعي الإسلامي، وثمة ما هو أعمق مما هو وارد في اتفاقية سيداو يتميز به ذلك القانون على حد زعم القول في التكشيف السياقي.

وفي موضوع الحجاب والنقاب، فكل الدراسات تشير إلى أنه ليس خيارًا للمرأة، مع وورد كشاف يتناقض في السياق ذاته، وهو أن الحجاب مفروض على الزوجة وليس ملزما للخادمة في المنزل وخارجه، وتكررت الوحدة المعجمية «فتنة» كمسوغ استراتيجي خطابي لضرورة الحجاب. أما موضوع سفر المرأة بلا محرم، فثمة اتجاه إجماعي حول منعه إلا بموافقة ولي الأمر (الأب أو الأخ «الوكيل الشرعي»). وورد ٧٦ كشافا سياقيا ممتدا يتضمن عدة أنماط خطابية متداخلة: تنقّل المرأة محدود بوجود المحرم، وتنقّل السائقين الخاصين مع المرأة مقبول، واجتماع الرجال والنساء في بيئة عمل مفتوحة محرمة، وفصل الأطباء والطبيبات في المستشفيات (النمط الأخير متعلق بالصحف السعودية).

أما في الجدول (٨)، فإن معظم أنماط الخطاب العام حملت نصائح وإرشادات، أو حملت وقائع غير ذات صلة بالنسوية واللانسوية، أو وردت الموضوعات على شكل تعداد مع موضوعات أخرى غير ذات صلة بالتحليل، أو سيقت في استطرادات عن وحدة الخطاب دون أي ارتباط بين الاستطراد والنص الوارد فيه بالنسوية واللانسوية.

الموضوع	التكرار	الحالات الإيجابية	الأنماط الخطابية
الزواج	٢٤٠٦	١٤٠٨	خيار مفتوح
الطلاق	٥٨١	١٢٥	خيار مفتوح
الخلع	١٤٣	١٢٩	تأييد الخلع بلا أسباب/ تأييده وإن كان على المظهر أو رائحة الرجل أو البيت
الحضانة	١٣٥	٦	من دون شروط
العنف الأسري	٢٨٧٠	١١٣١	ضد العنف/ مطالبة بقوانين
التعليم	١٦٧٤	١٦١١	تعليم مفتوح/ دون قيود



التمييز ضد المرأة	١٣٤٨	٢٣	تطبيق اتفاقية سيداو
الحجاب/ النقاب	٨٧٤	٧/ صفر	تزيين الحجاب/ ألوان الحجاب/ تطريز الحجاب
الرياضة النسائية	٤٥٩	٣	مبررات حجبها غير صحيحة/ غير علمية
السفر من دون محرم	٢٦٦	٢٣٠	مقبول في العصر الحالي/ توفر الأمان/ إمكانية الحج بدون رجل
الوصي (الرجل)	٨٦	١٠	- غير ضروري في الأماكن العامة - حاجته في سياق انعدم الأمان
قيادة السيارة	٨٤	٢	النساء أفضل من الرجال في القيادة
التحرش الجنسي	٤٢	٣٣	مطالبات بسن قانون التحرش
نزع ولاية الرجل	٣	صفر	عند منع الرجل من حقوقها القانونية

الجدول (٦): تكرارات الحالات الإيجابية للموضوعات

الموضوع	التكرار	الحالات السلبية	الأنماط الخطابية
الزواج	٢٤٠٦	١٢٦	تعدد الزوجات دون قيد/ خيار مقيد
الطلاق	٥٨١	١٠٤	ليس خيارًا للزوجة/ تكافؤ النسب
الخلع	١٤٣	١٤	الخلع غير مقيد
الحضانة	١٣٥	١١	مشروطة بصلاحية الأم
العنف الأسري	٢٨٧٠	٨٥٤	مسموح فيه العنف الناعم soft power



الموضوع	التكرار	الحالات السلبية	الأنماط الخطابية
التعليم	١٦٧٤	٦٣	تعليم مقيد بما يناسب وضع المرأة الجسدية والاجتماعية
التمييز ضد المرأة	١٣٤٨	١٠١٢	لم تعالج/ غير متقيدة بالأنظمة الدولية
الحجاب/ النقاب	٨٧٤	٨٦٧	مفروض/ عدم إثارة غرائز الرجل/ تركه ليس خيارًا
الرياضة النسائية	٤٥٩	١٦	مضر بوظائف جسم المرأة الحيوية
السفر من دون محرم	٢٦٦	٣٦	محرّم/ عدم إمكانية الحج بدون محرم
الوصي (الرجل)	٨٦	٧٦	ضروري/ استثناء السائق الأجنبي/ إرضاع البالغ/ انقطاع/ الفصل بين الأطباء والطبيبات/ حرمة رؤية الزوج أو الزوجة من الطرفين بعد الوفاة
قيادة السيارة	٨٤	١٣	محرمة عليها/ محرمة مع السائق الأجنبي/ تتطلب قوة جسمية في حال تعطل المركبة
التحرش الجنسي	٤٢	٩	بعض أنواع الحجاب يسبب التحرش/ التحرش سببه المرأة/
نزع ولاية الرجل	٣	٣	سبب لتصدع الأسرة القبائلية/ العشائرية

الجدول (٧): تكرارات الحالات السلبية للموضوعات



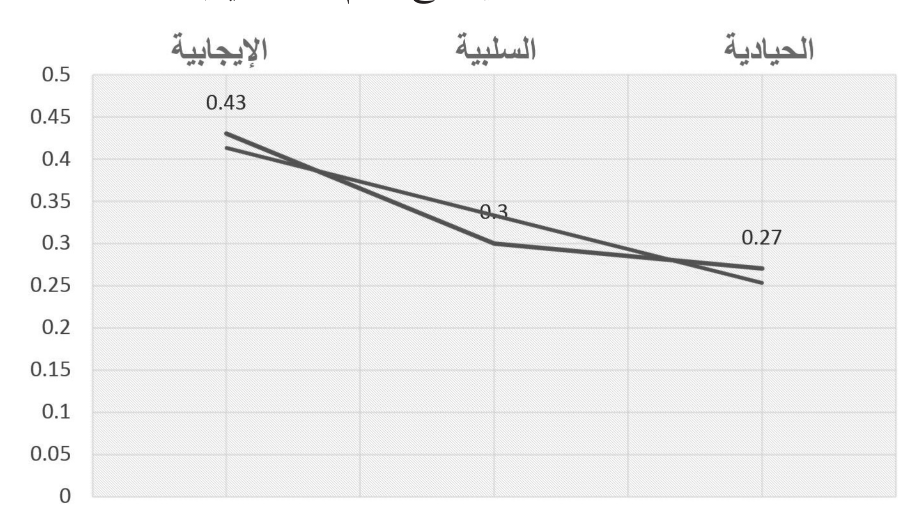
الموضوع	التكرار	الحالات الحيادية	الأنماط الخطابية
الزواج	٢٤٠٦	٨٧٢	نصائح/ الزواج من غير المسلمة
الطلاق	٥٨١	٢١٨	معدود ضمن عدة موضوعات غير ذات صلة
الخلع	١٤٣	١٣٣	معدود ضمن عدة موضوعات غير ذات صلة
الحضانة	١٣٥	١١٨	معدود ضمن عدة موضوعات غير ذات صلة
العنف الأسري	٢٨٧٠	٨٨٥	معدود ضمن عدة موضوعات غير ذات صلة
التعليم	١٦٧٤	صفر	لا شيء
التمييز ضد المرأة	١٣٤٨	٣١٣	معدود ضمن عدة موضوعات غير ذات صلة
الحجاب/ النقاب	٨٧٤	صفر	لا شيء
الرياضة النسائية	٤٥٩	٤٤٠	معدود ضمن عدة موضوعات غير ذات صلة
السفر من دون محرم	٢٦٦	صفر	لا شيء
الوصي (الرجل)	٨٦	صفر	لا شيء
قيادة السيارة	٨٤	صفر	لا شيء
التحرش الجنسي	٤٢	صفر	لا شيء
نزع ولاية الرجل	٣	صفر	لا شيء

الجدول (٨): تكرارات الحالات غير الإيجابية وغير السلبية للموضوعات



وعلى الرغم من وجود تكرارات لبعض الموضوعات المحددة بمحمل النسوية تفوق عدد تكرارات الموضوع نفسه الذي حمل اللانسوية، وعلى الرغم من وجود العكس أيضا لبعض الموضوعات كذلك، إلا أن المجمل في النتائج الضخمة بجمعها معًا تشير إلى رجحان الإيجابية على السلبية بمعدل بسيط (الشكل ٤). وهنا سر دقة مناهج التحليل الخطابي والتحليل الخطابي النقدي وفق منهج الفوخطابي على النصوص الضخمة المتنوعة ببنياتها النصية ووحداتها الخطابية.

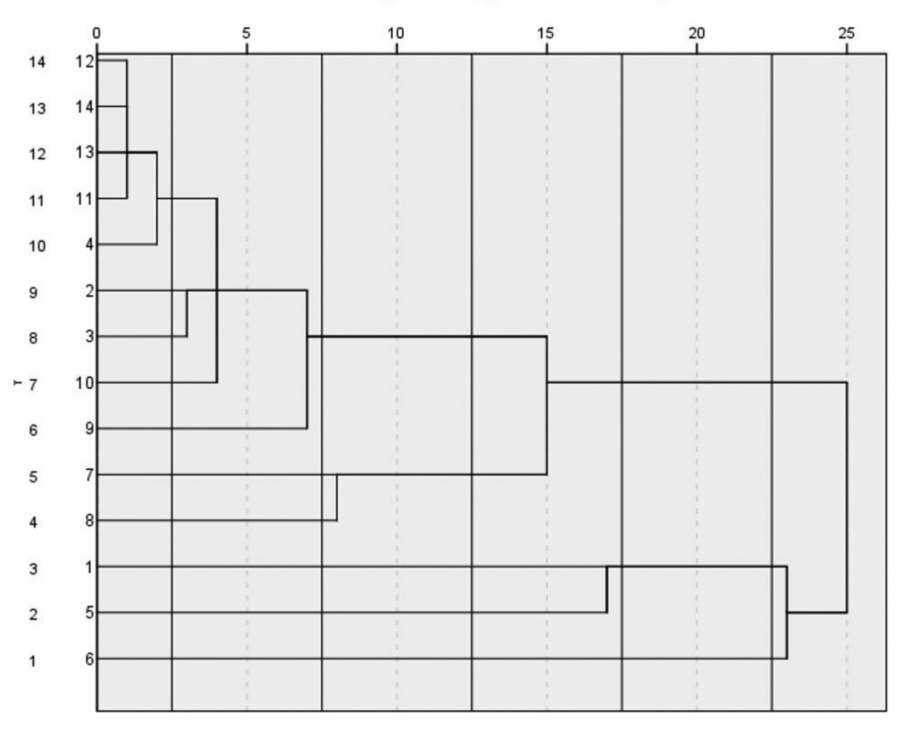
ولا يقتصر الأمر في تحليل مدونة الدراسة على النسوية واللانسوية، إذ من الممكن أن يوسع تحليل الخطاب أو تحليل الخطاب النقدي بشكل أكبر عن الموضوعات المحللة: كالتركيز على أصناف القواعد والوحدات المعجمية في نصوص المدونة المتعلقة بتشكل خطاب الممارسات الاجتماعية الخاص بالمرأة والنساء، أو التركيز على طبيعة الفاعلين وواقع الخطاب في تشكل السلطة والهيمنة وفقا لأنواع الإحالات النصية وغير النصية، أو التركيز على المجموعات الدلالية semantic preferences والنظم الدلالي semantic prosody برصد مجموعة حقول دلالية على مستوى الكلمات، أو التركيز على النظم الخطابي discourse prosody برصد ما يتصاحب مع كلمة محددة من كلمات معجمية دالة تُفسّر وتُشرح بناء على السياق الخطابي، مثل كلمة «الفتنة» وما يتصاحب معها من كلمات معجمية سياقية خطابية تتعلق بها في نسيج النظم الفوخطابي في مدونة الدراسة.



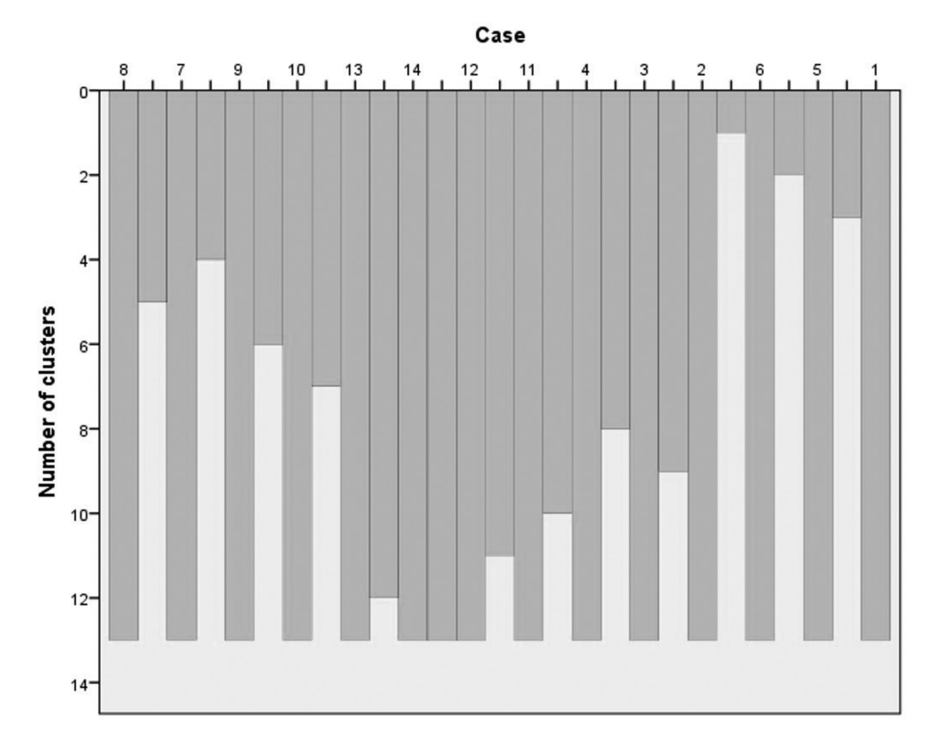
الشكل (٤): قيمة الكسر من ٠,٥ للإيجابية ٠,٤٣ وللسلبية ٠,٣٠ وللحيادية أو غير ذات الصلة ٠,٢٧ لمدونة الدراسة الخاصة بموضوع التحليل.

وعليه فإن طبيعة إعادة بناء الفوخطابي بمعالجة الموضوعات المحللة وفق تحليل العَنْقدة الهرمية hierarchal clustering التي تجلي طبيعة العلاقة الانضوائية لجسم مدونة الدراسة بشكل مصنف ومعاد ترتيبه مهمٌّ جدًا، وكمية المتّجهات vectors التي تتضمن كل تكرار خاص بموضع معين قد عولجت إحصائيا لتبيان متوسط كل متغير رقمي (تكرار) مع المتغير الاسمي (الحالة أو الموضوع) في الشكل (٥). وقد استعملتُ قياس التوسّط العنقودي centroid من أجل قياس موقع كل عنقود مع كل موضوع، ومتوسط المتجهات عن التوسط مقابل متوسط متجهاته vectors بين تكرارات بقية موضوعات التحليل (العناقيد). وفائدة هذا المنهج تكمن في فتح أفق أوسع للتحليل على البيانات الضخمة، كما أنه يعيد ترتيب الوحدات الخطابية الداخلية، ويعيد ترتيب البنيات النصية داخل كل خطاب، ويعيد كذلك ترتيب الأنماط الخطابية مهما كانت طبيعة التحليل على مستوى التواصلية الخطابية القائمة على مناهج الـ BCU؛ وذلك كله من أجل فهم أوسع لطبيعة التحليل التصاعدي لموضوعات الخطابات العربية الطبيعية.





الشكل (٥): عنقود مُتّجهات vectors كل متغير اسمي (كل موضوع) مع بقية المتغيرات الاسمية بحسب أرقامها في الجداول (٦)، و(٧)، و(٨) عن تكرار الحالات.



الشكل (٦): عدد العناقيد مقابل الحالات بحسب أرقامها في الجداول (٦)، و(٧)، و(٨) عن تكرار الحالات..

ولو عولج لنصيب حجم الوحدات الخطابية (الأخبار، والمقالات، والتقارير) كلُّ وحدة بذاتها لتغيرت نتائج التحليل لما هو دون الفوخطابي metadiscourse؛ كأن يكون التحليل في مجال الأخبار دون المقالات والتقارير، وهنا تكمن مشكلة تحليل الخطاب وتحليل الخطاب النقدي القائم على بيانات صغرى أو موضوعات صغرى أو ربما- أيضا-مجالات domains صغرى. وعليه فإن لمنهج التحليل الفوخطابي رؤية أوسع من نتائج التحليل الخطابي الخاص بحدث معين وبخطاب محدد ومناسبة ضيقة المقام.



١ . ٩ . الخاتمة

تنطلق الموضوعات التحليلية الخطابية من الوحدات اللغوية النصية السياقية إلى ما وراء تلك الوحدات من فاعلين وأثر وتواصلية خطابية متطابقة أو متقاربة أو متباينة، كونها–أي: الموضوعات– تفاصيلَ التحليل الفوخطابي metadiscourse. وأفاد هذا الفصل في تحليل خطاب النسوية واللانسوية (مع/ أو ضد؛ أي: الأصوات الإيجابية والأصوات السلبية الكمية وغيرهما) جزءًا من التطبيق دون تبيان التفاصيل الفاعلة، وذلك بمناهج الـ BCU. وعلى خلاف ما تقدمه الكشافات السياقية من أدوات معالجة اللغة الطبيعية للكلمة المركزية (المرأة) و(النساء) مثلا، تطرق هذا الفصل إلى طرائق تحليل الصورة الأولية للمنهجية المعتمدة على مناهج (بايبر وكونر وأبتون Biber, Connor and Upton 2007) حول موضوعات التحليل: الزواج، والطلاق، والخلع، والحضانة، والعنف، والتعليم، والتمييز، والحجاب، والرياضة، والسفر، وولاية الرجل، وقيادة السيارة، ونزع ولاية الرجل. وبالخلوص إلى أهم النتائج، أنبه القارئ إلى أهم التطبيقات من أجل التوسع التحليلي الخطابي القائم على الفوخطابي، وهي على النحو الآتي:

أولاً: إن المعالجة من الوحدات الخطابية إلى الموضوعات (كما هو مطبق في هذا الفصل) ومن ثمة من الموضوعات إلى الوحدات يعطي فهما أوسع لطبيعة الفوخطابي metadiscourse.

ثانيًا: إن هذه المناهج مبكرة–حتى في الأبحاث العالمية–؛ لكنها ستُعدّ الاتجاه السائد في العقود القادمة من القرن الحادي والعشرين.

ثالثًا: واقع تحليل الخطاب بمناهج بايبر وكونر وأبتون Biber, Connor and Upton (٢٠٠٧)، وبالاعتماد على أدوات وأساليب المعالجة الآلية مهما كانت مصادرها الموثوق بها تحليلا ومنهجيًا (وهي عديدة وكثيرة)، يشير إلى أن الاتجاه إلى النصوص الضخمة، ومعالجة أنماط لغوية وحجاجية وخطابية وتواصلية ضمن عدد ضخم من الوحدات الخطابية المتضمنة في بنية نصية واحدة، وضمن عدد ضخم من البنى النصية المتضمنة



في مدونة التحليل الخطابي أو التحليل الخطابي النقدي، قد بات ملحًا؛ لأنها (أي: هذه الاتجاهات) تفيد البحث والتحليل للخطاب المتعدد الأوعية والمجالات بعدة نتائج دقيقة وشاملة، نذكرها على النحو الآتي:

الأولى: موضوعات تحليل الخطاب، التي جرت العادة فيها بأدوات تحليل الخطاب التقليدية وذلك بالانطلاق من الإجابة عن الأسئلة الآتية: كيف تُحلّل، ولماذا، دون السؤال: ماذا تحلل؟ التي تشجعها مناهج الـ BCU، ولكن، بالاعتماد على البيانات الضخمة.

الثانية: يوظَّف هذا المنهج على البيانات اللغوية الضخمة، وبأي منهج في تحليل الخطاب أو تحليل الخطاب النقدي شريطة أن يحقق شرط «تواصلية الخطابات» (الفوخطابي)، ولا يمكن إجراؤه على بيانات لغوية صغيرة الحجم.



قائمة المراجع

أولا: المراجع العربية

- خطابي، محمد. لسانيات النص، مدخل إلى انسجام الخطاب. المركز الثقافي العربي، بيروت، ٢٠٠٦، الطبعة الثانية.
- شارودو، باتريك ومنغو، دومينيك. معجم تحليل الخطاب. ترجمة: عبدالقاهر المهيري وحمّادي صمود. تونس: دار سيناترا، ٢٠٠٨م.
- ماكنري، توني وهاردي، أندرو. (٢٠١٦). لغويات المدونة الحاسوبية: المنهج والنظرية والتطبيق. ترجمة سلطان بن ناصر المجيول. دار جامعة الملك سعود، الرياض.
- مجموعة مؤلفين. (١٩٩٦). الوضع القانوني والاجتماعي: دراسات ميدانية في ثمانية بلدان عربية مع دراسة تأليفية. المعهد العربي لحقوق الإنسان، تونس. مسترد في الرابع والعشرين من شهر أكتوبر، ٢٠١٦م، من http://www.aihr-iadh.net/pdf/etudes/Femme_arabe.pdf.
- المجيول، سلطان بن ناصر. البحث اللغوي في المدونات العربية الحاسوبية بين الممكن والمحتمل والمأمول. في: المدونات اللغوية العربية بناؤها وطرائق الإفادة منها. تحرير: صالح العصيمي. مركز الملك عبدالله الدولي لخدمة اللغة العربية، الرياض، ٢٠١٥م.

ثانيا: المراجع الأجنبية

- Alfaifi, A. & Atwell, E. (2016). Comparative evaluation of tools for Arabic corpora search and analysis. International Journal of Speech Technology, 9(2): 347-357.
- Al-Thubaity, A. (2015). A 700M+ Arabic corpus: KACST Arabic corpus design and construction. Language Resources and Evaluation, 49(3), pp. 721-751.
- Al-Thubaity, A., & Al-Mazrua, M. (2014a). Khawas: Arabic Corpora Processing Tool USER GUIDE. Retrieved October 2, 2016, from http://www.sourceforge.net/projects/kacst-acptool/files/?source=navbar.
- Al-Thubaity, A., Khan, M., Al-Mazrua, M., & Almoussa, M.



(2014b). KACST Arabic Corpora Processing Tool ‘‘Khawas’’ [Computer Software]. Retrieved October 2, 2016, from https://sourceforge.net/projects/ghawwasv4/.

Baker, P. (2006) Using Corpora in Discourse Analysis. London: Continuum.

Baker, P. and McEnery, T. (2005) ‘A Corpus-based Approach to Discourses of Refugees and Asylum Seekers in UN and Newspaper Texts’, Journal of Language and Politics 4(2): 197–226.

Baker, P., and Ellece, S. (2011). Key Terms in Discourse Analysis. London: Continuum.

Baker, P., Gabrielatos, C., Khosravinik, M., Krzyzanowski, M., McEnery, A., and Wodak, R. (2008) A useful methodological synergy? Combining critical discourse analysis and corpus linguistics to examine discourses of refugees and asylum seekers in the UK press. Discourse and Society, 19(3): 273-306.

Biber, D. (1995). On the role of computational, statistical, and interpretive techniques in multidimensional analyses of register variation: a reply to Watson. Text 15(3): 341–70.

Biber, D. Connor, U. and Upton, T. (2007). Discourse on the Move: Using corpus Analysis to Describe Discourse Structure. Amsterdam: John Benjamins.

Burr, V. (1995). An Introduction to Social Constructionism. London: Routledge.

Connell, R. (1987). Gender and Power. Stanford: Stanford University Press.

Fairclough, N. (1989) Language and Power. London: Longman.

Fairclough, N. and Wodak, R. (1997) ‘Critical Discourse Analysis’. In: van Dijk, T. A. (ed.) Discourse as Social Interaction, pp. 258–84. London: SAGE.

Gramsci, A. (1985). Selections from the Cultural Writings 1921–1926. In: Forgacs, D. and Nowell Smith, G. [Translated by W. Boelhower]. London: Lawrence and Wishart.



Kennedy, G. (1998). An introduction to Corpus Linguistics. London: Longman.

Offenhauer, P. (2005). Women in Islamic Societies: A Selected Review of Social Scientific Literature. A Report Prepared by the Federal Research Division in the Library of Congress under an interagency Agreement with the Office of the Director of National Intelligence/ National Intelligence Council (ODNI/ADDNIA/NIC) and Central Intelligence Agency/Directorate of Science and Technology. Retrieved July 14, 2016, from <https://www.loc.gov/rr/frd/pdf-files/Women_Islamic_Societies.pdf>.

Orpin, D. (2005) 'Corpus Linguistics and Critical Discourse Analysis: Examining the Ideology of Sleaze', International Journal of Corpus Linguistics 10(1): 37–61.

Ottaway, M. (2004). Women's Rights and Democracy in the Arab World. Massachusetts: Carnegie Endowment for International Peace.

Paltridge, B. (2007). Discourse Analysis: An introduction. London: Continuum, 1st edition.

The World Bank (2009). The Status and Progress of Women in the Middle East and North Africa. Washington DC: The Social and Economic Development Group and the Office of the Chief Economist of the Middle East and North Africa. Retrieved October 14, 2016, from <http://siteresources.worldbank.org/INTMENA/Resources/MENA_Gender_Compendium-2009-1.pdf>.

Tognini-Bonelli, E. (2004). Working with corpora: Issues and insights. In: Coffin, C. Hwings, a. and O'Lalloran, K., (eds.), Applying English Grammar: Functional and Corpus Approaches. London: Arnold, pp. 11-24.

Tohmé-Tabet, A. (2001). Women, Intangible Heritage and Development in the Arab World. International editorial meeting and future activities in the domain Iran-National Commission for UNESCO, Tehran June 2001. Retrieved October 24, 2016, from <http://www.unesco.org/culture/ich/doc/src/00161-EN.pdf>.



- Upton, T. A. and Cohen, M. A. (2009). An Approach to Corpus-based Discourse Analysis: The Move Analysis as Example. Discourse Studies, 11(5): 585-605.

- van Dijk, T. A. (1988). News as Discourse. Hillsdale (New Jersey): Erlbaum.

- van Dijk, T. A. (1993). «Principles of critical discourse analysis». Discourse and Society, London. Newbury Park and New Deli, 4(2), pp. 249-283.

- van Dijk, T. A. (1998) Ideology. London: SAGE.


هذه الطبعة
إهداء من المركز
ولايسمح بنشرها ورقياً
أو تداولها تجارياً

الفصل الثاني

د. محمود عثمان الحاج

المعالجة الآلية للعربية الإدارية: دراسة مدونات الإدارة وأخبار الأسهم والاقتصاد في الوطن العربي

الملخص

يعرض هذا الفصل طريقة بناء المدونات الصغيرة، والمعروفة بالمدونات المختصة. ويقف على ذكر ثلاث طرق أساسية: الويب بوصفه مدونة، واستخدام مدونات أخرى، وتقشير شبكة المعلومات. ويركز باستعمال تلك الطرق على تحليل اللغة العربية الإدارية المنتجة من رئيس مجلس الإدارة والرئيس التنفيذي مع المساهمين من ذكور وإناث. وقارنا بين العربيات الإدارية المستخدمة في المدونات الثلاث عن طريق استخدام أدوات تحليل المدونة الحاسوبية. وأظهرت التحليلات المبدئية في هذا الفصل عن وجود خلل وفجوة في تواصل وخطاب الإدارة مع المساهمين، وكيف أن المساهمين يتجهون إلى مصادر أخرى كالأخبار الاقتصادية وأخبار الأسهم لمعرفة الوضع المالي للشركة. واستنتج الفصل بإجراء التجارب على تصنيف وثائق المدونات آليا باستخدام التعلم الآلي بأن الآلة–وباستخدام بيانات تدريب جيدة–تستطيع نسخ التصنيف اليدوي من قبل الخبير المدوّني الحاسوبي. كما يخلص الفصل إلى أهمية لغويات المدونة الحاسوبية في دراسة لغة ما، حيث يوفر هذا الحقل بالأدوات والبرامج المعدة مسبقا كثيرًا من الوقت والجهد الذي يضيع وقت الباحث.



الكلمات المفتاحية: اللغة العربية الإدارية، بناء المدونات، مقروئية النص، معالجة المدونات، تصنيف الوثائق آليًا.

٢ . ١ . المقدمة

يُعدّ الإفصاح من المواضيع الحيوية والمهمة التي تلقى اهتماما كبيرا في السنوات الأخيرة. يعبر الإفصاح في التقارير المالية عن المعلومات الضرورية التي تكفل الأداء للشركات. ويلزم مبدأُ الإفصاح الكامل الشركاتِ بتوفير المعلومات اللازمة لمن اعتادوا على قراءة التقارير المالية لاتخاذ قرارات مستنيرة بشأنها. ومؤخرا زاد الاهتمام وبشكل ملحوظ بالإفصاح المحاسبي والاتصال المباشر بين الإدارة والمساهمين. ويتم ذلك باتصال مباشر عن طريق الكلمات أو الحروف التي تلخص حقائق وأرقام يصعب على المساهم الإلمام بها. وتُعدّ مرحلة الإفصاح من أهم المراحل التي تمكن الإدارة من توفير معلومات وحقائق حول موضوع معين، ويتم ذلك عن طريق إصدار تقارير مالية سنوية أو نصف سنوية تخاطب وبشكل مباشر المساهمين.

وتناقش هذه التقارير الأوضاع الحالية والمستقبلية للشركة بالإضافة إلى المتغيرات والتطورات التي قد تؤثر على الوضع المستقبلي للشركة باستخدام لغة منمقة ومقنعة منحازة. ويكون سرد هذه الحقائق والمعلومات عن طريق خطاب مباشر من رئيس مجلس الإدارة–وهو أعلى منصب–لأي مجموعة منظمة لشركة ما. ويُعيّنُ الشخص عادة في هذا المنصب عن طريق الترشيح أو التعيين من أعضاء مجلس الإدارة، ويدير رئيس مجلس الإدارة الاجتماعات، وينظم عملها بطريقة مباشرة. وكما أسلفنا سابقا؛ يكون رئيس مجلس الإدارة هو المسؤول الأول عن الخطاب الموجه للمساهمين في الشركة، ويلي ذلك خطاب من الرئيس التنفيذي للشركة. ويشكل منصب الرئيس التنفيذي أعلى سلطة في الإدارة العليا بصفته المسؤول الأول والأخير عن إدارة الشركة أمام مجلس الإدارة. ويتضمن خطاب الرئيس التنفيذي سردًا لأداء الشركة خلال العام الماضي. ويقدم الخطابُ تفاصيلَ الإيرادات والنفقات عن طريق نظرة مالية عامة، ويشتمل كذلك على أهم التطورات المؤثرة على الشركة، بالإضافة إلى الأهداف والغايات للسنة القادمة. وبشكل عام: يُعدّ التقرير السنوي الذي يحتوي على خطابات رئيس مجلس الإدارة، والرئيس التنفيذي وسيلةً لإعطاء تقرير عن حالة الشركة للمستثمرين والمساهمين أو أي طرف ذي اهتمام مرتبط بأداء الشركة. ولا يقتصر مخاطبة المساهِمين على



الشركات والإدارة فقط، بل يتعدى ذلك قانونيًا إشراك وكالات الأخبار الاقتصادية وأخبار الأسهم. وتقوم وكالات الأخبار بدور كبير في توضيح وتبيان ما تعنيه خطابات الشركات والجهات الإدارية، ويشتمل ذلك على إظهار ما قد تحتويه هذه الخطابات على رسائل ضمنية يصعب على المساهم فهمها أحيانا.

٢. ٢. الغرض من الدراسة

سألقي في هذا الفصل الضوء على أهمية اللغة العربية المستخدمة في كتابة خطابات رئيس مجلس الإدارة والرئيس التنفيذي، والمدرجة في التقارير المالية للشركات في مختلف القطاعات في الوطن العربي، مع مقارنتها بالتقارير الإخبارية المتعلقة بالوضع الاقتصادي للشركات وأخبار الأسهم. وتُعدّ التقارير المالية-أو بعبارة أخرى التقارير السنوية-المنتج النهائي الذي يصدر في نهاية السنة أو الفترة المالية للنظام المحاسبي، ويشتمل على معلومات مالية وغير مالية، وتُعدّان إحدى وسائل إيصال المعلومات للأطراف ذات العلاقة.

وفي هذا الفصل، سأهتم بالجزء اللغوي للتقارير السنوية الذي يسمى «الإفصاح»، وأسلّط الضوء على اللغة المستخدمة في كتابة خطابات إدارة الشركة، ومدى اختلافها أو تشابهها مع اللغة المستخدمة في كتابة أخبار الأسهم والأخبار الاقتصادية، وذلك عن طريق إجراء تجارب وتحليلات لغوية باستخدام أدوات حقل لغويات المدونة الحاسوبية ومناهجه.

٢. ٣. بناء المدونة

هناك العديد من المعايير المستخدمة لبناء المدونات من التي تتمثل في حجم المدونة ومدى تنوع اللغة المستخدمة، بالإضافة إلى خصائص النص: كعمر الكاتب وجنسه. وتختلف المعايير بينها، غير أنها تشترك في المجال الذي تُستخدم به المدونة، والفترة الزمنية المغطاة، وطريقة استخدام هذه المدونة بشكل منهجي مقنن. وطريقة (سنكلير Sinclair 1991) في بناء المدونة عامة، وتشتمل على جمع أكبر قدر ممكن من النصوص من مختلف المجالات، وبمحاولة تُحيط بأكبر قدر ممكن من تنوع اللغة. والمثال على ذلك: بناء مدونة للغة العربية الفصحى المكتوبة والمحكية.

ولكن، ومع ازدياد مستخدمي المدونات العامة وتطبيقاتها بشكل ملحوظ، زاد



الاحتياج إلى مدونات أصغر مهتمة بجزء معين من اللغة، أو بنوع لغوي آخر وعلى نطاق ضيق. وأطلق(بارتينغتون Partington 2004) على هذه المدونات الصغيرة مصطلح «المدونات المختصة أو أحادية النوع» (انظر أيضا بارتينغتون Partington 2003).

ولهذه الدراسة غاية نهايتها إنشاء مدونة مختصة باللغة العربية الطبيعية المنتجة فعلا في المجال الإداري والاقتصادي. وتشمل هذه المدونة على ثلاثة أنواع رئيسية: خطابات الإدارة، وأخبار الأسهم والأخبار الاقتصادية. ويتضمن النوع الأول على تقارير الرئيس التنفيذي ورئيس مجلس الإدارة من مجموعة كبيرة من الشركات في مختلف أقطار الوطن العربي. أما النوع الثاني والثالث؛ فيحتويان على أخبار محلية ودولية عن أسعار الأسهم والأحداث المرتبطة بتذبذب الأسعار، وعن أخبار اقتصادية بشكل عام قد تؤثر سلباَ أو إيجابا على أسعار الأسهم.

٢ . ٤ . طريقة بناء المدونة

جعل كونُ المدونة متضمنةً اللغة المكتوبة جمعَها وترتيبها أسهل فيما لو كانت محكية؛ لأن النص يحتاج إلى تدقيق من مختص في اللغة للتأكد من صحة التدوين في حال نسخ الكلام المحكي.

وجمع المدونة وبنائها؛ قمنا باستخدام ثلاثة طرق هي على النحو الآتي:

أولا: الويب بوصفه مدونة (Web-as-corpus).

ثانيًا: استخدام مدونات أخرى.

ثالثًا: تقشير شبكة المعلومات (Web scraping).

٢ . ٤ . ١ . الويب بوصفه مدونة

اعتمدتُ هذه الطريقة لجمع أكبر عدد ممكن من تقارير الرئيس التنفيذي ورئيس مجلس الإدارة. وتعتمد طريقة الويب كمدونة على استخدام الشبكة المعلوماتية لبناء مدونة مختصة (Kilgarriff 2001) متماشيا مع تعريف المدونة بكونها مجموعة من النصوص (McEnery and Wilson 1996). وبناءً على هذا الأساس وهذا التعريف، قمتُ بجمع تقارير من الويب عن طريق كتابة برنامج حاسوبي يقوم بالبحث وتحميل التقارير أوتوماتيكيا عن طريق البحث في محرك بحث جوجل بإستخدام جُمل قمنا



باختيارها يدويا. الجمل المستخدمة هي: «كلمة رئيس مجلس الإدارة» و«كلمة الرئيس التنفيذي». واستخدمتُ مرادفات وعبارات أخرى تهدف إلى مساعدة البرنامج في الوصول إلى أكبر عدد ممكن من التقارير، على سبيل المثال: استخدام المؤنث «كلمة» مع «رئيس» و«تنفيذي» بالإضافة إلى استخدام «تقرير» بديلا عن كلمة «كلمة» والبحث وفقا لهذه التراتبية. وقمتُ بعد ذلك بفلترة يدوية للتقارير التي تم تحميلها أوتوماتيكيا للتأكد من كونها تقارير من رئيس مجلس الإدارة والرئيس التنفيذي. وأجريتُ تنظيفًا للتقارير بشكل يدوي، وذلك بإبقاء النص المتعلق بالتحليل، وإزالة أي نص أو محتوى لا يمت للتقرير بصلة.

وحمّلَ البرنامجُ أكثر من ألف تقرير، وقلصتُ هذا الرقم إلى ٤٠٠ تقرير من شركات في الخليج العربي، وبلاد الشام، وشمال إفريقيا وذلك بالفلترة اليدوية. وتضمنت المحصلة النهائية تقارير رئيس مجلس الإدارة مشكلة نصف مدونة الويب، وتقارير الرئيس التنفيذي التي شكلت النصف الثاني من مدونة الويب.

٢. ٤. ٢. استخدام مدونات أخرى

قمنا في بحث سابق بإنشاء مدونة حاسوبية أسميناها «كلمات» (El-Haj and Koulali 2013)، وتختص هذه المدونة في النصوص الإخبارية بمجموعة مختلفة من الأقسام ومن ضمنها الأخبار الاقتصادية. وقمنا باختيار ٤٠٠ مقالة عشوائية من أجل مساواة عدد التقارير التي تم جمعها بالطريقة السابقة.

٢. ٤. ٣. تقشير شبكة المعلومات (Web scraping)

يزخم الويب بالبيانات والمعلومات، ولكن أكثر هذه البيانات غير مقروءة للآلة كما هي للإنسان. وقد تظهر بشكل ملائم ومنسق للقارئ غير أنها تُخزن في هيئة تجعل من استخدامها للفهرسة الأوتوماتيكية أو في المدونات الحاسوبية أمرًا صعبًا، ويعود ذلك لطريقة عرض النص على الموقع. ولو أخذنا على سبيل المثال موقعًا إخباريًا، سنجد أن الخبر يظهر بشكل واضح في منتصف الصفحة، غير أن نظرة سريعة في زوايا الموقع تكشف ما يحيط بالنص من عناوين وصور وروابط لمقالات أخرى ناهيك عن روابط الموقع، والدعايات، وتعليقات القراء في أحيان أخرى. ولو أننا بصدد بناء مدونة مختصة بالأخبار عن طريق جمع آلاف المقالات من مواقع إخبارية مختلفة، فسيكون من



الصعب إتمام العمل بنسخ النص من كل مقال بشكل يدوي. ولتسريع هذه العملية: قمنا باستخدام طريقة تقشير الويب مع مراعاة الخصوصية وحقوق الطبع. ويتم ذلك عن طريق بناء برنامج حاسوبي بسيط يقوم بمحاكاة المستخدم في عرضه ونسخه للنص مما يساعد على توفير الوقت والجهد اللازم لتحميل النصوص.

وفي هذا الجزء، قمنا ببناء مدونة صغيرة مكونة من ٤٠٠ مقال تهتم بأسعار الأسهم المحلية والدولية وأحوالها من موقع investing.com. وبنينا برنامجًا حاسوبيًا يتصفح الموقع ويُحمّل المقالات الإخبارية من قسم أخبار أسواق الأسهم. ويستخلص البرنامج بعد ذلك النص من صفحة المقال عن طريق إزالة لغة الترميز وإبقاء النص. كما دربنا البرنامج على استخلاص الجزء المتعلق بالخبر والتخلص من باقي النصوص في كل صفحة.

٢. ٥. التوسيم الآلي للمدونات

باستخدام الطرق الثلاث الوارد ذكرها في المبحث (٢. ٤.) أصبح حجم المدونة النهائية ١٢٠٠ مقال إخباري وتقارير إدارية، جميعها مكتوبة بلغة عربية فصيحة (أو نموذجية). وتحتوي المقالات على نصوص مكتوبة على شكل نص خام بدون ترميز وبدون هيكلة. ويتفاوت كتّاب المقالات ما بين إناث وذكور مكتوبة بلغة إدارية منمقة. وتحتوي مدونة الخطابات الإدارية على أكثر من ١٢٣ ألف كلمة، بينما تتكون مدونة الأخبار الاقتصادية من ١٨٢ ألف كلمة، أما مدونة أخبار الأسهم فقد بلغ حجمها ١١٤ ألف كلمة.

ولتسهيل العمل على تحليل المدونات التي تم بناؤها سابقا قمنا باستخدام أداة التوسيم الآلي في جامعة ستانفورد[1] التي تقوم بالعمل مباشرة على النص العربي، وأعدتُ هذه الأداة بواسطة تركيب تقنيات القواعد اللغوية والقواعد الإحصائية والرموز التي تم بناؤها على توصيف النحو التقليدي للغة العربية. وتنقسم الترميزات إلى عدة أقسام؛ منها ثلاث رئيسة وهي الأسماء والأفعال والأدوات، أما الظروف وحروف الجر فتُعامل كجزء من الأجزاء الرئيسة. ووُضعت الترميزات على الكلمة كاملة مع زوائدها. وتم عمل مجزئ للكلمة مبني على مُرمز صرفي ومُرمّز إحصائي.

١ – انظر: http://nlp.stanford.edu/software/tagger.shtml



ووظيفة المجزئ هي رفع الزوائد وتحديد جذر الكلمة ويتم هذا في مرحلة أخرى. وبما أن الباب النحوي للكلمة يحَدَد عن طريق الزوائد فإن الرموز تحدد عن طريق المجزئ. وقد ثبتت دقته بنسبة ٥ ,٩٦ بالمائة، وقد تم تدريب الأداة باستخدام مدونة تري-بانك العربية (Arabic Tree-bank p1-3). وقد استخدمتُ الأداة لتوسيم المدونات الثلاث الوارد ذكرها في المبحث (٢. ٤).

٢. ٦. أدوات تحليل مدونة الدراسة

وظفت لتحليل المدونات ومقارنة النصوص واللغة المستخدمة عدة أدوات مجانية وبرامج متوفرة وبشكل مباشر للقارئ، وأسرد بشكل مبسط هذه الأدوات قبل الخوض في تفاصيل كل منها لاحقا، وقبل عرض نتائج التحليل والمقارنة بين أقسام المدونة.

٢. ٦. ١. مقروئية النص العربي

يُعدّ حساب مقروئية النص وسيلة لقياس مدى سهولة قراءة نص معين. ويتم ذلك عن طريق حساب تكرار الكلمات والجمل في النص، بالإضافة إلى عدد مقاطع الأصوات. وقد يشمل ذلك أيضا عدد الكلمات الصعبة والكلمات الطويلة. ويطلق مصطلح «كلمات صعبة» على الكلمات التي يتجاوز طولها عدد معين من الأحرف. فمثلا في اللغة السويدية وفي أغلب المعادلات المستخدمة لحساب مقروئية النص تُعرّف الكلمات المحتوية على أكثر من ٦ أحرف بـ «الكلمات الطويلة» (Björnsson 1968) بينما يطلق على الكلمات المحتوية على أكثر من ٣ مقاطع صوتية بـ «الكلمات المعقدة» (Gunning 1968).

ويعود حساب الكلمات الصعبة والمعقدة إلى الاعتقاد السائد بأن صغار السن وحديثي اللغة يحجمون عن استخدام هذه الكلمات لكونها لم تتوغل في قاموس كلماتهم، ولذلك يعد النص المتضمِّن بكثرة على كلمات معقدة وصعبة بأنه نص صعب القراءة، وتُعرّف صعوبة القراءة هنا مثلا بكون النص مقروءًا من قِبَلِ قرّاء ذوي درجات علمية متقدمة، والعكس صحيح، إذ اُعتبر النص السهل قراءةً في حال كان قراؤه في مراحل دراسية ابتدائية، والمثال على ذلك: قُرّاء الصف الأول الى الصف السادس ابتدائي.

ولتوضيح ذلك يقوم بعض المعادلات الموضوعة لحساب مدى مقروئية النص على وضع نظام تقدير العمر أو المرحلة الدراسية للشخص القادر على قراءة النص المراد حسابه.



على سبيل المثال: تقوم معادلة فليش (Kincaid et al., 1975) الموضحة أدناه بحساب مقروئية النص بإصدار ناتج رقمي يقع بين الصفر والمائة. ويبين الجدول (١) الترجمة النصية والمرحلة الدراسية والأعمار المناسبة لهذه الأرقام، وهو ما يُعتدّ به في وصف النص بكونه سهلاً أو صعبَ القراءة:

فليش= ٢٠٦,٨٣٥ – ١,٠١٥ (عدد الكلمات/ عدد الجمل) – ٨٤,٦ (عدد مقاطع الأصوات/ عدد الكلمات)

النتيجة	المرحلة الدراسية	الترجمة النصية
٩٠,٠ ١٠٠,٠	الصف الخامس	سهل جدا للقراءة. يسهل فهم النص من قبل طالب يبلغ من العمر ١١ عاما
٨٠,٠ ٩٠,٠	الصف السادس	سهل القراءة
٧٠,٠ ٨٠,٠	الصف السابع	سهل القراءة إلى حد ما
٦٠,٠ ٧٠,٠	الصف الثامن إلى التاسع	إنجليزية صافية يسهل قراءتها من قبل طلاب بعمر ١٣ إلى ١٥ عاما
٥٠,٠ ٦٠,٠	الصف العاشر إلى الحادي عشر	صعب القراءة إلى حد ما
٣٠,٠ ٥٠,٠	في المرحلة جامعية	من الصعب قراءة النص
٠,٠ ٣٠,٠	حامل لدرجة جامعية	صعب جدا. يفهم من قبل خريجي الجامعات

الجدول (١): الترجمة النصية لمعادلة فليش.

وتقوم هذه المعادلات بتوفير الوقت والجهد اللازم لحساب مقروئية النص حيث تستبدل هذه الأرقام الحاجة إلى عدد كبير من القرّاء اللازم إيجادهم لتحديد مقروئية النص.

وكون معادلة فليش تعمل بشكل صحيح للغة الإنجليزية فقط قمنا بحساب مقروئية النص للمدونة عن طريق استخدام معادلة «عثمان» (El-Haj and Rayson 2016) التي تحسب مدى سهولة قراءة النص بطريقة مشابهة لمعادلة فليش. وتم حياكة



معادلة عثمان للعمل مع النصوص العربية فقط، ويتم ذلك عن طريق حساب أعداد الكلمات والجمل بالإضافة إلى الكلمات الصعبة والمعقدة عن طريق عد الحروف والمقاطع الصوتية. وتوضح المعادلة التالية كيفية حساب معادلة عثمان التي بدورها تنتج رقمًا بين الصفر والمائة -كما في الجدول السابق-لحساب معادلة فليش:

عثمان = ٢٠٠,٧٩١ - ١,٠١٥ × (عدد الكلمات / عدد الجمل) - ٢٤,١٨١ × (عدد الكلمات الصعبة / عدد الكلمات + عدد المقاطع الصوتية / عدد الكلمات + عدد الكلمات المعقدة / عدد الكلمات + عدد الأحرف الفصحى / عدد الكلمات)

وتحتوي الكلمات الصعبة هنا على أكثر من ٥ أحرف، بينما تحتوي الكلمات المعقدة على أكثر من ٤ مقاطع صوتية تم الوصول إليها عن طريق تحليل مليوني كلمة في مدونة تشكيلة (El-Haj and Rayson 2016) التي استخدمت لتمرين معادلة عثمان. ولتمييز النص العربي عن غيره يُعوّل على المعادلة في حساب عدد الأحرف العربية، وهي حروف معينة تم اختيارها لكونها تهمل من قبل الكتّاب ضعيفي اللغة والإملاء، وهذه الحروف هي: «ء»، «ئ»، «وء»، «ذ»، «ظ»). وسنوضح لاحقا نتائج مقروئية النص للمدونة عن طريق مقارنة مقروئية النص للأقسام المختلفة في المدونة.

٢. ٦. ٢. غواص ACPTs[١]

وهي أداة لمعالجة المدونات العربية تم تطويرها في مدينة الملك عبدالعزيز للعلوم والتقنية، ويقوم البرنامج بحساب تكرار الكلمات والتكرار النسبي والتوافق اللفظي بالإضافة إلى الكلمات الأكثر ارتباطا. ويقوم البرنامج أيضا بمساعدة اللغويين العرب بشكل أساسي في دراسة اللغة اعتمادا على البيانات والإحصاءات التي توفرها المدونات عن طريق الأداة. وسأوظّف هذه الأداة في دراسة اللغة المستخدمة في التقارير السنوية، والأخبار المالية والاقتصادية؛ من أجل التعرف على أكثر الكلمات ارتباطا، وذلك باستخدام الكشاف السياقي والمتصاحبات اللغوية التي قد تساعدنا في فهم طريقة التصريح في هذه المدونات المختصة باللغة العربية الإدارية.

١- انظر: https://sourceforge.net/projects/ghawwasv4 (الثبيتي وآخرون Al-Thubaity et al. ٢٠١٤ أ، ٢٠١٤ب).



٢. ٦. ٣. آنت-كونك (AntConc)

وهو برنامج آخر لمعالجة المدونات. ويتميز آنت-كونك بشهرته وكثرة استخدامه في مجال معالجة المدونات وخصوصا اللغة الإنجليزية (Anthony 2014). وسنقوم باستخدام آنت-كونك بشكل مباشر لمعالجة المدونات التي تم توسيمها آليا كما ذكرنا سابقا، وذلك باستخدام الموسّم الآلي من جامعة ستانفورد. ويقرأ آنت-كونك الرموز وحساب تكرارها وتصاحبها، الأمر الذي سيساعد على معرفة الرموز المستخدمة في اللغة العربية الإدارية.

٢. ٧. معالجة المدونات الإدارية

أسردُ في هذا الجزء التجارب والتحليلات الناتجة عن استخدام أدوات تحليل المدونات السالف ذكرها. ونحاول في هذا الجزء فهم المدونات العربية الإدارية ومعرفة أسلوب السرد المستخدم في المؤسسات ونشرات الأخبار الاقتصادية والمالية، وسيعين ذلك على فهم الجزء الناقص في خطابات الإدارة والنشرات مع المساهمين.

٢. ٧. ١. مقروئية النص

نحسب بداية مقروئية النص باستخدام معادلة عثمان لحساب مقروئية النص العربي. ونسرد النتائج بالتفصيل لكل مدونة مع شرح وتحليل النتائج. وكما أسلفنا سابقا، تقوم معادلة عثمان بحساب مقروئية النص عن طريق إعطاء قيمة تقع بين الصفر والمائة. ولتبيان مقروئية المدونات الثلاث التي تم جمعها، قمنا بحساب المعادلة لكل مقال وخطاب ابتداء بخطابات رئيس مجلس الإدارة والرئيس التنفيذي. وتوضح الأرقام في الجدول (٢) المتوسط الكلي لمقروئية النص في المدونة، بالإضافة إلى متوسط الكلمات الصعبة والمعقدة، وعدد الكلمات والجمل ومقاطع الأصوات.

المتوسط الرياضي	مدونة الخطابات الإدارية	مدونة الأخبار الاقتصادية	مدونة أخبار الأسهم (المالية)
معادلة عثمان	٤٩,٤١	٣٣,٢٢	٨٨,١٠
الكلمات الصعبة	١٢٤,٥٠	١٤٥,٤٩	٧١,٧٥



المتوسط الرياضي	مدونة الخطابات الإدارية	مدونة الأخبار الاقتصادية	مدونة أخبار الأسهم (المالية)
الكلمات المعقدة	٧٠,٤٠	٩٢,٤٥	٣٠,٧
الكلمات الفصيحة	١٣	١٤,٧٣	٨,٩٣
الكلمات	٣١٩	٤١٧,٠١	٣٣٣,٣٩
الجمل	١١	١٢,٢١	١٣,٥٦
المقاطع الصوتية	١٣٣٢	١٧١٦	٨٣٦,٧٢

الجدول (٢): مقروئية المدونات باستخدام معادلة عثمان.

وبالنظر إلى نتائج معادلة عثمان والمتوسطات الحسابية في الجدول (٢) نرى أن مدونة أخبار الأسهم أسهل قراءة من المدونات الأخرى. ويظهر ذلك بالنظر إلى الوسط الحسابي لناتج معادلة عثمان، الذي تخطّى الـ ٨٠ نقطة. وإذا قمنا بترجمة الرقم نصيا كما في الجدول (١) سنرى أن النص يسهل قراءته من لدن طلاب المستوى الدراسي السادس. وتشير المتوسطات الحسابية إلى قلة استخدام الكلمات الصعبة والمعقدة مقارنة بالمدونات الأخرى، إلى قلة استخدام أقل للكلمات الفصيحة التي تم تعريفها في (El-Haj and Rayson 2016) على أنها الكلمات المحتوية على أحد الحروف الآتية: (ء، ئ، ؤ، ذ، ظ) أو المنتهية بـ (ون، وا) شريطة أن تحتوي الكلمة على أكثر من أربع مقاطع صوتية. ويدل هذا على أن الكُتّابَ قد اتجهوا إلى استخدام كلمات بسيطة وكفيلة بإيصال المعلومة للمساهمين بشكل أسرع. قد تكون هذه الكلمات المختارة نتاج ترجمة من لغة أخرى إلى العربية، وهذا نعلله بعدم الحاجة أصلا إلى لغة منمقة ومنقحة من أجل إيصال معلومة عن ارتفاع سهم ما أو انخفاضه للمساهمين.

وعلى عكس ذلك نرى أن مدونة الخطابات الإدارية أصعب في القراءة من تلك المستخدمة لكتابة أخبار الأسهم، ويحاول كل من رئيس مجلس الإدارة والرئيس التنفيذي في الخطابات الإدارية بمخاطبة المساهمين بشكل مباشر مع أخذ الحيطة والحذر عند الكتابة بألا تشير أي من الكتابات إلى عدم استقرار أو انخفاض في إيرادات الشركة.



لذلك نرى أن الخطابات منمقة ومقنعة ومكتوبة بلغة لبقة توحي للقارئ باستقرار الشركة واستمراريتها. لذلك نلاحظ ازديادًا في استخدام الكلمات المعقدة والصعبة، وهي الكلمات التي يتجاوز عدد المقاطع الصوتية فيها عن ٤ أو يزيد عدد حروفها عن ٥ (يعرض الجدول ٣ بعضًا من الكلمات الصعبة والمعقدة المستخدمة في خطابات رئيس مجلس الإدارة والرئيس التنفيذي).

الاتصالات	الركائز	المجتمعات
المعلومات	توقعاتهم	الاستثمارات
مواجهتها	بجهودكم	إستراتيجي
الاستيعابية	بالإنجازات	الإمكانيات
مقترحاتكم	الديناميكية	الأكاديميون

الجدول (٣): كلمات صعبة ومعقدة في خطابات الإدارة.

وبالنظر إلى الجدول (٢) مرة أخرى نرى أن الأخبار الاقتصادية أصعب في القراءة من أخبار الأسهم و خطابات الإدارة، ويعود ذلك إلى كثرة استخدام الكلمات الصعبة والمعقدة في كتابة هذه المقالات، وقد يُبرر ذلك بطبيعة خبرة من يصوغون التقارير واتساع مصطلحاتهم حيث نرى انخفاض قيمة معادلة عثمان لدرجة أنه يصعب على من ليس في مرحلة جامعية أن يقرأها بسلاسة، ولذلك نجد أخبارًا كهذه موجهة إلى رجال الأعمال وأصحاب الشركات والسياسيين في بعض الأحيان حيث إن الأخبار الاقتصادية تساعد في فهم الحياة اليومية للمواطن وكذلك وضع الأسواق العالمية بالنسبة للشركات.

ويحتوي الجدول (٤) على بعض الكلمات المتقدمة معجميًا والمستخدمة في صياغة الأخبار الاقتصادية، ومع أنها تتشابه في طولها وصعوبتها مع تلك المستخدمة في خطابات مجلس الإدارة والرئيس التنفيذي إلا أنها مستخدمة بشكل أكبر في كتابة الأخبار الاقتصادية مقارنة مع الخطابات الإدارية.

للبتروكيماويات	مايكروسوفت	الاستشارات
المشروبات	الاستهلاكية	المعدنيات
الاحتياطيات	الانخفاضات	الإسترليني
الارتفاعات	الكهربائية	الإماراتية
الرئيسيتين	التداولات	الإيجارات

الجدول (٤): كلمات صعبة ومعقدة في الأخبار الاقتصادية.

ولنلقِ الضوء على اللغة المستخدمة في صياغة خطابات مجلس الإدارة والرئيس التنفيذي التي تندرج تحت مسمى الإفصاح المحاسب، حيث نجد بالنظر إلى النتائج أن اللغة المستخدمة في الصياغة وفي مخاطبة المساهمين لغةٌ صعبة بشكل عام، في الوقت الذي يفضل فيه المساهمين عادة مقالات أسهل لمعرفة وضع الشركة واستقرارها في المستقبل بشكل مباشر. ولذلك تلعب أخبار الأسهم دورا كبيرا في إيصال المعلومات المرادة للمساهمين مباشرة دون الحاجة إلى الخوض في تفاصيل خطابات الإدارة. ويدل هذا على وجود خلل في اتصال الشركات بالمساهمين حيث إن اللغة المستخدمة في خطابات الإدارة غير مقروئية رغم كون وظيفتها الأساسية المحاولةَ في تلخيص الحقائق والأرقام المحاسبية التي يصعب على المساهمين فهمها.

وهنا نرى كيف يساعدنا تحليل المدونة على اكتشاف هذا الخلل عن طريق معادلات وإحصائيات بسيطة تشير إلى أهمية المعالجة الآلية للمدونات حيث إنها ستساعدنا في كشف معلومات لا تظهر بشكل واضح لعين القارئ أحياناً. وسأدرس في المبحث القادم نص المدونات من أجل محاولة لفهم هذا الإشكال عن طريق دراسة وتحليل بعض الأمثلة.



٢. ٧. ٢. معالجة المدونات

أُعالج في هذا الجزء مدونات الدراسة وذلك بدراسة نصوصها وبتحليل الكلمات والترميزات المستخدمة فيها، إلى جانب دراسة التصاحب اللفظي ودراسة اللغة بشكل أعمق. وكما أسلفنا سابقا سنقوم بمعالجة المدونات باستخدام أدوات غوَاص وآنت-كونك مع شرح النتائج والتحليلات. وأبتدئ التحليل باستخدام الكشاف السياقي في أداة غواص من أجل دراسة استخدام بعض الضمائر المنفصلة، ومن أجل فهم طريقة الخطاب سواء كان من قبل الإدارة أو المقالات الإخبارية.

٢. ٧. ٣. حساب تكرار الكلمات

بمعالجة المدونات الثلاث بأداة الغواص لغرض حساب تكرار الكلمات، يعرض الجدول (٥) الكلمات العشرين الأكثر تكرار. ويوضح الجدل (٥) تكرار بعض الكلمات مرتبة تنازليا، ويُلحظ بأن الكلمات الشائعة تطغى على الاستخدام في جميع المدونات بينما تتميز كل مدونة بكلمة أو كلمتين مرتبطة ارتباطا كليا بنوع نصوص المدونة. ويُلحظ أيضًا تكرار كلمات مثل قبيل: مجلس، والإدارة ورئيس، وذلك في المدونة الإدارية، أما في مدونة أخبار الأسهم، فيكثر استخدام كلمات الأسهم، والسهم، والسعر.

ترتيب التكرار	المدونة الإدارية		مدونة الأخبار الاقتصادية		مدونة أخبار الأسهم	
	الكلمة	التكرار	الكلمة	التكرار	الكلمة	التكرار
١	خلال	٦٠٨	في	٥١٨٧	عند	٤٢٥٦
٢	مع	٥٨٩	من	٤٧٥٠	في	٤٠٤٦
٣	مجلس	٥٥٥	على	٢٥٠٤	او	٢٤٤٢
٤	الشركة	٥٤٧	ان	٢٣٩٢	سهم	٢٠٩٢
٥	هذا	٤٨٩	الى	٢٠١٥	نقطة	١٧٢٤



ترتيب التكرار	المدونة الإدارية			مدونة الأخبار الاقتصادية			مدونة أخبار الأسهم	
	الكلمة	التكرار		الكلمة	التكرار		الكلمة	التكرار
٦	عن	٤٥٥		التي	١١٢٢		على	١٦٤٤
٧	الادارة	٤١٩		خلال	٨٣٥		سعر	١٤٦٣
٨	الله	٤١٠		في	٨٢٩		الأسهم	١٤٦٢
٩	هذه	٤٠٢		هذه	٨١٥		نهاية	١٢٣٥
١٠	شركة	٣٩٥		عن	٧٤٢		بنسبة	١١٤٩
١١	الذي	٣٦١		هذا	٦٩٤		من	٩٨٦
١٢	ما	٣٦٠		مع	٦٧٣		الى	٩٨١
١٣	رئيس	٣٤٩		حيث	٦٥٣		مؤشر	٩٢٩
١٤	كما	٣٤٤		كما	٥٨١		مستويات	٧٩٠
١٥	حيث	٣١٥		بن	٥٣٠		شهر	٧٧٣
١٦	عام	٣٠٤		او	٤٦٠		اما	٧٣٨
١٧	العام	٣٠١		العمل	٤٥٢		بسعر	٧٣٥
١٨	العمل	٢٧٧		القطاع	٤٤٨		مستوى	٦٧٣
١٩	تحقيق	٢٧٥		العام	٤٣٩		واغلق	٦٥٣
٢٠	العربية	٢٧٠		ريال	٤٣٩		نحو	٦٣٢

الجدول (٥): تكرار الكلمات في المدونات الثلاث.



٢ . ٧ . ٤ . استخدام ضمائر المخاطبة

لو أخذنا على سبيل المثال الضمائر المنفصلة «نحن» و «أنا» لوجدنا أن رئيس مجلس الإدارة والرئيس التنفيذي يخاطبان المساهمين باستخدام (نحن) في معظم خطاباتهم في حين اختار ٣ رؤساء مجلس إدارة مخاطبة المساهمين باستخدام الضمير المنفصل (أنا). ويوضح الجدولان (٦) و(٧) استخدامهما خطابات مجلس الإدارة.

الجملة السابقة	الكلمة	الجملة اللاحقة
والمجتمعات التي نخدمها. وها	نحن	اليوم نواصل وضع معايير الريادة
وفكره في مسيرة نجاح الشركة.	نحن	نؤمن بأن سر نجاحنا في
آل نهيان «طيب الله ثراه».	نحن	في شركة الاتحاد للقطارات نقوم
وفكره في مسيرة نجاح الشركة.	نحن	نؤمن بأن سر نجاحنا في
المشاريع الزراعية والمشاريع الصغيرة والمتوسطة.	نحن	نحرص على تطوير قطاع المشاريع
دعائم داخلية وخارجية قوية، وها	نحن	نضع بين أيديكم التقرير السنوي
والنظرة المتوازنة لخططها المستقبلية، وها	نحن	نرفع هذه التوصية إلى جمعيتكم
استخدام المصطلح بشكل صريح. نعمل	نحن	هنا في بوبا العربية من
تاريخ دخول المرسوم حيز التنفيذ.	نحن	ملتزمون كشركة بتطوير نظم نقل

الجدول (٦): استخدام الضمير المنفصل نحن في خطابات الإدارة.

الجملة السابقة	الكلمة	الجملة اللاحقة
على دعمهم المستمر لنا	أنا	منبهر بمسيرة حتى الآن،
كما في نصف السنة الأول،	أنا	على ثقة من أننا ستتمكن

الجدول (٧): أمثلة على استخدام الضمير المنفصل أنا في خطابات الإدارة.



ويكمن الفرق بين الجدولين (٦) و(٧) في طريقة مخاطبة الإدارة للمساهمين، فغالبية رؤساء مجلس الإدارة والرؤساء التنفيذيين يتحدثون باسم الشركة مما يجعل الخطاب عامًا غير مخصص للقارئ الذي يشعر بأن الخطاب موجه من الشركة لا من الرئيس الإداري شخصيا.

ولفهم طريقة السرد نجد أن أخبار الأسهم تخلو تماما من استخدام أي ضمائر منفصلة كـ: (نحن) و(أنا) حيث إنها أخبار قصيرة مهتمة بتذبذب أسعار الأسهم وبالتغييرات الطارئة على السوق المالي، ونجد كذلك استخدامًا قليلاً للضمائر المنفصلة في الأخبار الاقتصادية، كما نجد استخدام هذه الضمائر مرتبطًا بإعادة ما ذكرته الإدارة في خطاباتها، مما يجعل هذه المقالات عامة للمساهمين، ولا تتحدث نيابة عن شركة مساهمة. أما الفروق البسيطة كاستخدام الضمائر المنفصلة فهي تساعدنا على فهم طرق الاتصال وما قد يساعد المساهم في فهم أحوال السوق المالي وسوق التداول.

وقد يعني هذا أن الأخبار الاقتصادية وأخبار الأسهم تقومان بتبسيط خطابات الإدارة، وإعادة نشرهما للمساهمين بشكل مبسط، مما يدل على أن خطابات مجلس الإدارة تعاني من ضعف في إيصال المعلومات المرادة للمساهمين.

٢. ٧. ٥. استخدام ضمائر الملكية وتغليب المذكر على المؤنث

يُنظر في الأمثلة التالية إلى استخدام ضمائر الملكية المذكرة والمؤنثة مع المقارنة بينها. ولو أخذنا على سبيل المثال الضمائر «لهم»، و»لهن» و»لكم» لوجدنا في جميع الخطابات استعمالَ الإدارة الضميرَ المذكر لـ «هم»، والضمير المحايد لـ «كم» في خطاباتهم مع المساهمين. إنه من الجدير بالذكر بأن التقارير التي تم جمعها وتمت كتابتها من قبل رؤساء مجلس إدارة ورؤساء تنفيذيين ذكور تقتصر على ٣ خطابات فقط مكتوبة من قبل رؤساء مجلس إدارة نساء. ويظهر أيضا عن طريق الأمثلة والكشافات السياقية بأن الإدارة تخاطب المساهمين بصيغة الجمع دون تحديد الجنس وذلك باستعمال: «لهم» و «لكم». وتخاطب الإدارة العملاء بالجمع المذكر الذي يشمل أيضا المؤنث كما في كلمة «المساهمين» حيث لا نجد أي تكرار لكلمة المساهمات في الخطابات. ويتكرر المثال عن طريق استعمال كلمات «إخوة» و»أخوات» حيث قام أربعة من رؤساء مجلس الإدارة والرؤساء التنفيذيين بمخاطبة المساهمات النساء باستعمال كلمة «أخوات»، بينما



فضلت الأغلبية على مخاطبة الذكور فقط باستعمال كلمة «إخوة» أو «الإخوة»، وتخطوا كلمات «إخواني» و»أخواتي» على حذو ما سبق. ويُوضح الجدول (٨) هذه الفروقات في الخطاب، واختيار الإدارة لمخاطبة الذكور، متناسية المساهِمات الإناث، حتى وإن كان ذلك ضمنيا، وتحليل النص ساعدنا هنا على فهم طريقة المخاطبة واستخدام اللغة العربية في الخطابات الإدارية. وقد لا يتعدى الأمر كونه تعميمًا، ولكن: قد يساعد ذلك على تسليط الضوء على الأسباب المتعلقة بعدم ذكر المساهمات ومخاطبتهن بشكل مباشر.

الجملة السابقة	الكلمة	الجملة اللاحقة
بإحسان إلى يوم الدين. السادة	المساهمين	الكرام، السلام عليكم ورحمة الله
السادة	المساهمين	الكرام، الحضور الأفاضل، السلام عليكم
ونسبة العائد على متوسط حقوق	المساهمين	إلى ٩, ١٥٪ ونسبة العائد على
وختاما أتوجه بالتحية والشكر لجميع	المساهمين	والعملاء الكرام على ثقتهم بنا
السادة	المساهمين	الكرام تحديات ومؤشرات يطيب لي
إلى ١١٢,٦٧٨,٠١٣ د.ك. ارتفاع حقوق	المساهمين	بنسبة ٤, ٦٪، لتصل إلى
أفضل الصلاة التسليم. إخواني وأخواتي	المساهمين	الكرام،،، بالأصالة عن نفسي وبالنيابة
إلا أن أتقدم بالنيابة عن	إخواني	أعضاء مجلس إدارة المجموعة وبالأصالة
آله وصحبه أفضل الصلاة التسليم.	إخواني	وأخواتي المساهمين الكرام،،، بالأصالة عن
عن	إخواني	أعضاء مجلس الإدارة، يسعدني أن
إلا أن أتقدم بالنيابة عن	إخواني	أعضاء مجلس إدارة المجموعة وبالأصالة
آله وصحبه أفضل الصلاة التسليم.	إخواني	وأخواتي المساهمين الكرام،،، بالأصالة عن
عن	إخواني	أعضاء مجلس الإدارة، يسعدني أن



الجملة السابقة	الكلمة	الجملة اللاحقة
بالإضافة إلى البيانات المالية المجمعة.	إخواني	المساهمين: استطاعت الراية العالمية العقارية
ديون مستحقة لأي جهة.	إخواني	المساهمين: لا يخفى عليكم الظروف
وانخفاضا مستمراً في الأسعار.	إخواني	المساهمين: بناء على هذه
أما بعد: فيسرني باسم	إخواني	وأخواتي أعضاء مجلس الإدارة والهيئة
أصالة عن نفسي ونيابة عن	إخواني	أعضاء مجلس الإدارة وجميع المساهمين
مكملين مسيرة النجاح والعطاء».	إخواني	وأخواتي الكرام،،، يسرني أن أرحب بكم

الجدول (٨): تغليب المذكر على المؤنث في خطابات الإدارة.

٢. ٧. ٦. دراسة الإيجابية والسلبية في الخطابات الإدارية

لو أنعمنا النظر في مجموعة من خطابات مجلس الإدارة والرؤساء التنفيذيين لرأينا أنها إيجابية بشكل دائم، ويدل ذلك على محاولة الإدارة في بث الثقة إلى المساهمات والمساهمين وخلق صورة بناءة للشركة. ولو أخذنا الكلمات الإيجابية مثل: «أمل» و»ثقة» و»نجاح»، وقارناها مع كلمات سلبية مثل «فشل» و»تراجع» و»هبوط» لوجدنا أن الإدارة تحاول وبشكل كبير تجنب استخدام الكلمات السلبية في الوقت الذي تسيطر فيه الكلمات الإيجابية على الخطاب بشكل عام. يوضح الجدول (٩) إيجابية الخطاب الإداري من قبل رؤساء مجلس الإدارة والرؤساء التنفيذيين.

الجملة السابقة	الكلمة	الجملة اللاحقة
مؤكدة. حقيقة تمكنا من مواصلة تحقيق النمو	والنجاح	وهي شهادة على الروح التي لدينا والفلسفة
التأثير في اقتصاد المعرفة العالمي، إن استمرار	النجاح	والتقدم لدولنا يعتمد على الاستمرارية في



الجملة السابقة	الكلمة	الجملة اللاحقة
قناة للتواصل المباشر مع عملائنا وشركائنا في	النجاح.	تعد شركة الموارد للاستقدام من أوائل الشرك
النجاحات التي تحققت والبناء عليها بثقة وتفاؤل	وأمل.	يحق لنا أن نفخر بالمكانة التي وصلت
ذلك تباعاً تقديم خدمات الكترونية حديثة وجديدة	نأمل	من خلالها تقوية وتعزيز موقع مصرفكم التنافسي
الأثاث في مصر أحد أهم الصناعات التي	نأمل	فيها القيام بدور فعال في استيعاب المزيد
كما في نصف السنة الأول، أنا على	ثقة	من أننا ستتمكن من تحقيق جميع الأهداف
بشركتنا على امتداد عقدين من الزمن، هذه	الثقة	الغالية التي دفعتنا للمضي قدماً في مسيرة
يشكل دعما» لنا بمواصلة خططنا الإستراتيجية بكل	ثقة	وطموح بتوفيق من الله جل وعلا. لقد

الجدول (٩): إيجابية الخطاب الإداري.

ولو نظرنا في الكلمات السلبية المستخدمة-وهي قليلة جداً-لوجدنا غلبة استعمال «تراجع» على جميع الكلمات السلبية الأخرى؛ لكونها كلمة سلبية مبطنة تشير إلى هبوطٍ في الأرباح غير أنها توحي للقارئ بأن هذا التراجع مؤقت كما في الأمثلة التالية في خطابات الإدارة في الجدول (١٠).

الجملة السابقة	الكلمة	الجملة اللاحقة
ايرادات شركة الاهلي كابيتال بنسبة	تراجع	بلغت ٢٩ بالمائة من القيمة
الدولي عن الكويت الى ان	تراجع	اسعار في الاسواق العالمية قد
و١٥٪ لمؤشر كويت ١٥ مع	تراجع	حاد في مستويات السيولة مقارنة



الجملة السابقة	الكلمة	الجملة اللاحقة
سوق الكويت المالي، حيث	تراجع	المؤشر الاهم الذي يقيس اداء
بـــ ١٠٠ نقطه) كما	تراجع	المؤشر الوزني بنسبة ١٣٪ محققاً
النفط العالمية العامل الابرز في	تراجع	اداء اسواق المال العربية خلال
الشركة ووسط اجواء مالية واستثمارية	تراجع	فيها حجم التداول في سوق
سوق الكويت المالي، حيث	تراجع	المؤشر الاهم الذي يقيس اداء
بـــ ١٠٠ نقطه) كما	تراجع	المؤشر الوزني بنسبة ١٣٪ محققاً
النفط العالمية العامل الابرز في	تراجع	اداء اسواق المال العربية خلال
بعد عقد من النمو، عبر	تراجع	القطاعات الرئيسية، المُتمثل في تمويل
حجم الدين والذي نتج عنه	تراجع	الدين بشكل ملحوظ بلغت نسبته

الجدول (١٠): سلبية الخطاب الإداري.

عن طريق معاينة الجداول في الأعلى نجد أن كلمة تراجع يكثر استخدامها للإشارة إلى خلل خارج نطاق الشركة وكما في الأمثلة يعود التراجع دائما إلى السوق المالي والأسعار العالمية وهو ما تستخدمه الإدارة كمبرر لهبوط أسعار أسهمها. ونرى هنا دور الكتابة المنمقة في إبعاد السلبية عن الشركة ونسب أي خلل لعوامل خارجية. وقام المؤلف (El-Haj et al., 2016) بتوضيح استخدام الشركات في نسب العوامل الخارجية لأي خلل يؤثر على أرباح الشركة وذلك عن طريق دراسة إيجابية النص وسلبيته، وربط ذلك بالأحداث الخارجية والداخلية للشركة. ووجد البحث أن الشركات تكون إيجابية في حال وجود أرباح وتنسب ذلك لعوامل داخلية كجودة العمل وتفوق الإدارة. بينما تنسب الشركات الخسائر إلى عوامل خارجية كالكوارث الطبيعية أو تراجع أسعار السوق المالي وهذا ما نلاحظه في الجدول (١٠) حيث تقوم الإدارة بتوجيه القارئ لوجود خلل في السوق المالي والمؤشرات والقطاعات الرئيسية. ونلاحظ في المثال



الأخير في الجدول (١٠) أن كلمة تراجع استخدمت بطريقة إيجابية للتنويه بتراجع في الدين العام للشركة.

وتكررت كلمة تراجع فقط ١٧ مرة في خطابات مجلس الإدارة والرئيس التنفيذي وعددها ٤٠٠ في الوقت الذي تكررت فيه كلمة تراجع ٥٤٩ مرة في كل من الأخبار الاقتصادية وأخبار الأسهم التي توضح وبشكل صريح تراجع أسعار أسهم الشركات مع ذكر أسمائهم وهو ما تتجنبه الإدارة في خطاباتها. ويوضح الجدول (١١) بعض الأمثلة على استخدام كلمة تراجع في كل من المدونة الاقتصادية ومدونة الأسهم.

الجملة السابقة	الكلمة	الجملة اللاحقة
من صعود السوق الاوسع حيث	تراجع	سهم انجلو امريكان ٨,٤ بالمئة
الاغلاق. سهم الوطنية للتبريد (DU:TABR)	تراجع	بمقدار ١,٣٨٪ او ٠,٠٢٠ نقطة
في الجلستين السابقتين مكاسبها مع	تراجع	سهم مجموعة طلعت مصطفى ٥,٥
الاسلامية القابضة ١,٦٪، الى جانب	تراجع	قطر الوطني ٠,٨٪، كما انخفض

الجدول (١١) السلبية في الأخبار الاقتصادية وأخبار الأسهم.

٢. ٧. ٧. دراسة الأفعال والأسماء في المدونات

نستخدم أداة آنت-كونك لهذه الدراسة كونها تستطيع قراءة النصوص العربية الموسومة برموز أداة التوسيم الآلي في جامعة ستانفورد. وبعد تحليل المدونة الإدارية وجدتُ أن خطابات الإدارة بشكل عام تفرط في استخدام الأسماء مقارنة مع الأفعال، واُستخدم أكثر من ٤٠٠٠٠ اسم مقارنة بـ ٨٠٠٠ فعل في ٤٠٠ خطاب إداري بحجم ١٢٣٢٨٩ كلمة. وتشكل الأسماء أكثر من ٣٦٪ من الخطاب الإداري في الوقت الذي شكلت فيه الأفعال فقط ١١٪، وتتساوى مدونات الأخبار الاقتصادية والأسهم بعدد الأفعال والأسماء حيث تشكل الأسماء أكثر من ٨٠٪ من النص مع تزايد في استخدام الأفعال حيث وصل ذلك إلى ٢١٪ من النص لكل مدونة حيث تتكون كل منهما من



٤٠٠ مقال. وتوضح الأرقام بأن الإدارة أقل تواصلا مع المساهِم مقارنة مع المقالات الإخبارية. ويشير العدد الضئيل من الأفعال إلى قلة تفاعل الإدارة مع المساهمين حيث يكثر الحديث عن إنجازات الشركة وتطلعاتها المستقبلية دون الحديث عن الأفعال والأحداث والأزمنة. وقد يعود ذلك بالنفع على المساهم، لكن قلة التفاعل قد تدل على وجود فجوة في الخطاب بين الإدارة والمساهمين وهو ما يعوضه المساهم عن طريق قراءة الأخبار الاقتصادية وأخبار الأسهم.

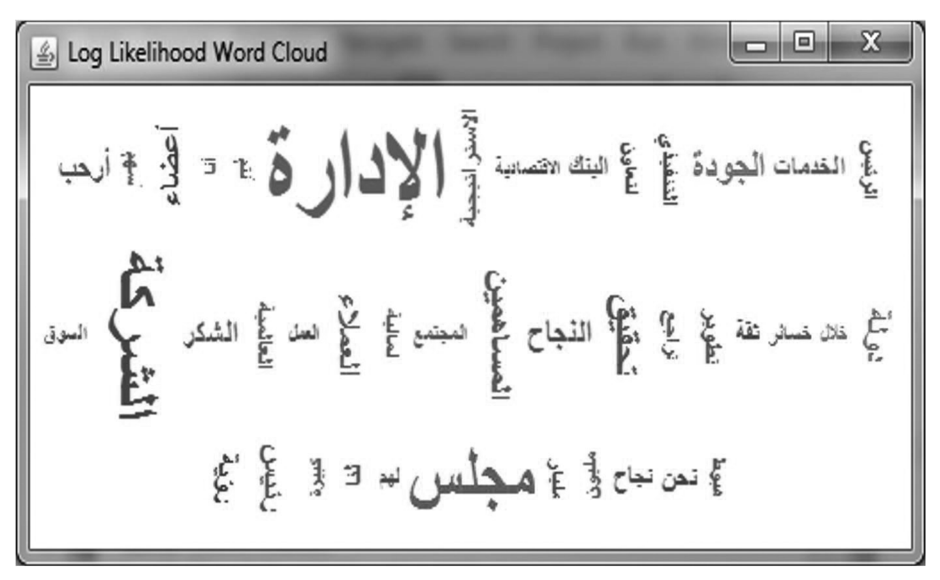
٢. ٨. سحابات الكلمات

ننشئ أخيرا سحابات الكلمات word clouds المتكررة بكثرة في مدونة الخطابات، ومقارنتها مع كل من مدونة الأخبار الاقتصادية ومدونة أخبار الأسهم. ولإنشاء السحابات قمنا باستخدام أداة سحابات كلمات معامل الاحتمال اللوغاريثمي[١] (Java_WordCloud_LogLikelihood). وتعين هذه الأداة على إنشاء سحابة كلمات معتمدة على تكرار الكلمات وعلى نتيجة معامل الاحتمال اللوغاريثمي الذي يشير إلى الكلمات ذات الأهمية الإحصائية، ويظهر ذلك عن طريق ملاحظة ازديادٍ في حجم الخط. ويوضح الشكل (١) مقارنة مدونة الخطابات الإدارية مع مدونة الأخبار الاقتصادية، ونرى بوضوح كيف أن كلمات مثل «الإدارة»، و»مجلس»، و»الشركة» هي أكثر الكلمات ذات الأهمية إحصائيًا، حيث يوضح حجم الخط أنها مستخدمة بشكل إحصائي أكبر من باقي الكلمات مثل: «الاقتصادية» المستخدمة بشكل أكبر في مدونة الأخبار الاقتصادية. ونرى أيضا كيف أن بعض الكلمات مثل «هبوط» و»خسائر» ضعيفة الدلالة التكرارية إحصائيا. ويتفق الشكل (٢) مع هذا عن طريق مقارنة مدونة الخطابات الإدارية مع مدونة أخبار الأسهم التي تشير إلى استخدام أكبر لكلمات الإدارة والكلمات الإيجابية، مقارنةً مع الكلمات السلبية أو حتى الحديث عن العملاء والمساهمين.

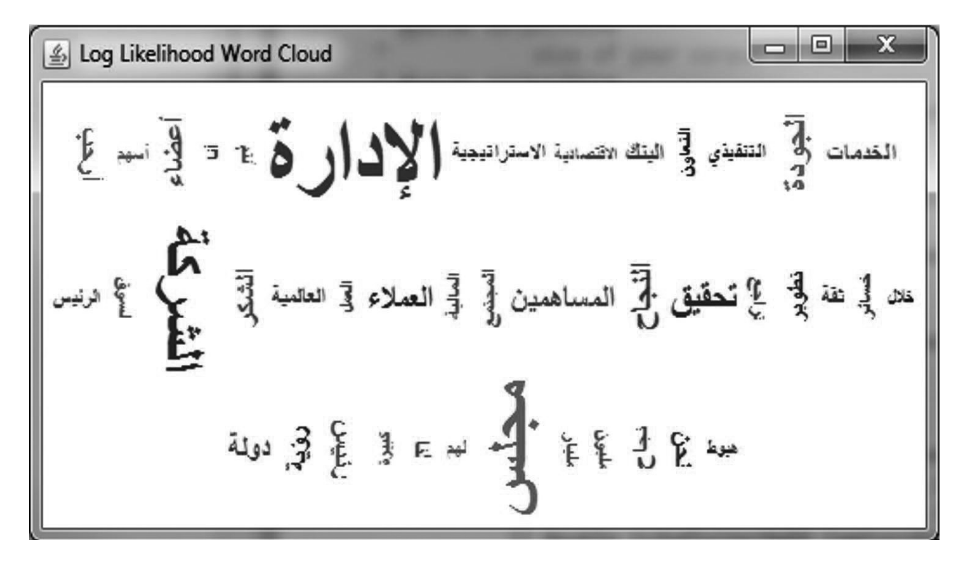
---

١ – انظر: https://github.com/drelhaj/Java_WordCloud_LogLikelihood





الشكل (١): مقارنة مدونة خطابات الإدارة مع مدونة الأخبار الاقتصادية.



الشكل (٢): مقارنة مدونة خطابات الإدارة مع مدونة أخبار الأسهم.

٢ . ٩ . تصنيف الوثائق باستخدام التعلم الآلي

التعلم الآلي هو أحد فروع الذكاء الاصطناعي الذي يتم عن طريقه تصميم وتطوير خوارزميات وتقنيات تسمح للحواسيب بتطوير خاصية التعلم. ولتلخيص ذلك يمكننا تعريف التعلم الآلي بأنه العلم والتقنية اللذين يسمحان للحاسوب بالتصرف من تلقاء نفسه دون الحاجة إلى برمجة مسبقة، ولا يعتمد الحاسوب على التلقين فقط ويعزو تصرفاته مما تعلمه في السابق؛ لأن تطور العلم والتقنيات الحاسوبية جعل التعلم الآلي مغايرا لما كان عليه في الماضي حيث أصبح أقل تكلفة وأكثر سرعة.

والتعلم الآلي ذو أهمية كبيرة في مجال تحليل المدونات؛ لأنه يوفر كثيرًا من الوقت والجهد والعمل اليدوي الذي قد يتكلفه الباحثون والعاملون في مجال المدونات الحاسوبية.

ومن أكبر تطبيقات تعليم الآلة وأكثرها شيوعا عملية تصنيف الوثائق التي تتطلب وقتا كثيرا من القراءة والفهرسة اليدوية اللازمة للوصول إلى قرار تصنيف وثيقة مختارة تحت مسمى معين. ويعرّف (Sebastiani 2002) تصنيف الوثائق على أنه العملية التي يتم فيها تصنيف نص مكتوب بلغة طبيعية إلى فئات موضوعية (categories and themes) تم اختيارها مسبقا.

ولو أخذنا على سبيل المثال الخطابات والمقالات في المدونات التي قمنا ببنائها مسبقا، لافترضنا أننا بصدد تصنيف إحدى هذه الوثائق إلى واحد من التصنيفات الثلاثة التي لدينا كون الوثيقة خطابًا إداريًا أو مقالاً اقتصاديًا أو أحد أخبار الأسهم. ولحل هذه المشكلة، فإنه يتعين على المصنّف أن يقرأ الوثيقة أو جزءًا منها قبل أن يأخذ قرار التصنيف، وهذا يتطلب خبرة مسبقة بخصائص المصنفات الموجودة والفئات الموضوعية، وينبغي على القارئ أن يميز بين الخطاب الإداري أو المقال الإخباري ومعرفة نوع المقال الإخباري في حال تم تصنيفه تحت ذلك المسمى. وقد يستطيع قارئ ذو خبرة عالية بتصنيف عشرات المقالات في ساعات قليلة، ولو أخذنا في عين الاعتبار الحجم الحقيقي للوثائق في الشركات لوجدنا أن كمية الوثائق الموجودة في أي شركة أو مؤسسة إخبارية تتجاوز ما قد يستطيع تصنيفه إلى مجموعة كبيرة من القراءة في أسبوع واحد. والحل الجذري لذلك هو بذل بعض الوقت والجهد البسيط لتعليم آلة، لتقوم

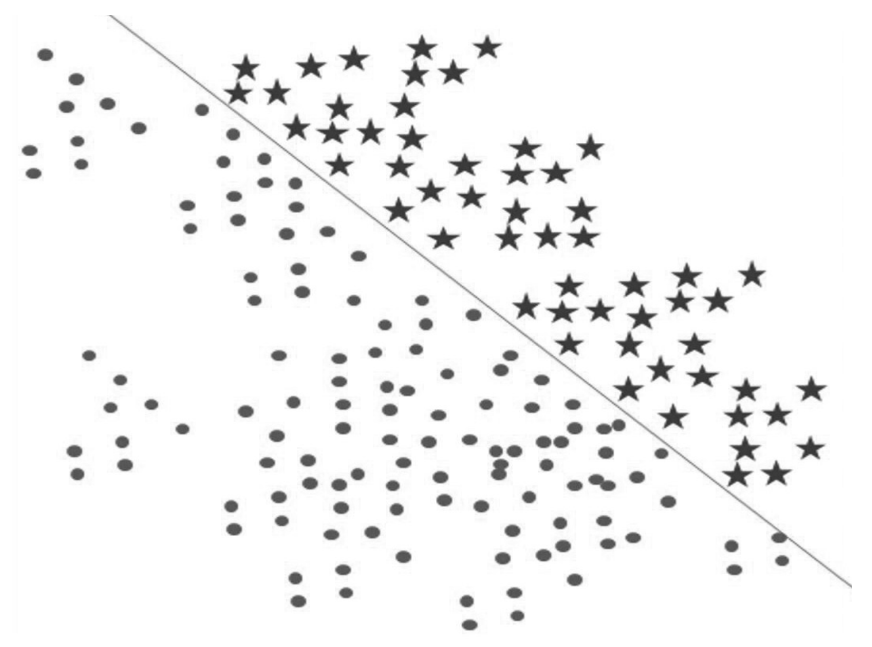


بدور المصنف كونه يساعد هذه الآلة على تصنيف الآلاف بل الملايين من الوثائق في مدة زمنية قد لا يتجاوز بضع دقائق.

وفي هذا الجزء نتبنى التعلم الآلي في تصنيف الوثائق التي لدينا عن طريق تعليم برنامج حاسوبي من أجل التفريق بين هذه الوثائق واختبار مدى جودة التعليم في تصنيف وثائق لم يرها أو يتعلمها من قبل.

ونستخدم في هذا الجزء تقنية التعليم الخاضع للإشراف Supervised Learning التي تسمى أيضا التعلم التنبئي Predictive Learning حيث تقوم الآلة بمحاولة التنبؤ بماهية المقال لتصنيفه إلى إحدى الفئات الموضوعية. وفي هذا النوع من التعلم يتم تدريب الآلة باستخدام مقالات تم تصنيفها مسبقا، ويتم ذلك عن طريق قراءة مقالات تم تصنيفها يدويا لغرض تعليم وتدريب الآلة، وتسمى هذه المدخلات باسم بيانات التدريب Training Data حيث تقوم الآلة باستخدام إحدى الخوارزميات لمحاولة تعلم الفرق بين هذه المقالات، ويتم ذلك عن طريق إيجاد أفضل الطرق للفصل بين المقالات. وتعتمد كفاءة الآلة على جودة بيانات التعلم، فكلما تقاربت المقالات في الشكل والمضمون صعب على الآلة التفريق بين المقالات. ولتوضيح ذلك: لنعتبر أن النجوم في الشكلين (٣) و(٤) ترمز إلى مقالات رياضية بينما ترمز الدوائر إلى مقالات سياسية، سنجد في الشكل (٣) بيانات التدريب تتكون من وثائق يسهل تصنيفها كون الفرق بين المقالات الرياضية والسياسية واضح، وتكفي الخوارزمية الخطية لها من أجل التفريق بينهما. أما في الشكل (٤)، فسنجد أن المقالات متداخلة مما يجعل استخدام الخوارزمية الخطية أمرًا صعبًا، ولذلك تتفوق الخوارزميات غير الخطية في تصنيف وثائق من هذا النوع. ونلحظ في الشكل (٤) أيضا أن الآلة صنفت بعض المقالات السياسية على أنها رياضية، وبالمثل صنفت بعض المقالات الرياضية على أنها سياسية، ويعود هذا إلى جودة بيانات التدريب، ونرى أن هنالك تقاربًا كبيرًا بين الوثائق. وتُعالج هذه المشكلة بمحاولة إيجاد مميزات features قد تساعد الخوارزمية على التمييز بين المقالات وهو ممكن كشفها وإيجادها بواسطة لغويات المدونة الحاسوبية.





الشكل (٣): تصنيف الوثائق باستخدام خوارزمية خطية.



الشكل (٤) تصنيف الوثائق باستخدام خوارزمية غير خطية.



وأبدأ تصنيف المدونات الثلاث بتحديد بيانات التعلم وبيانات الاختبار. ويفضل دائما لتعليم الآلة أخذ ٩٠٪ من نصوص المدونات لاستخدامها كبيانات تدريب وإبقاء ١٠٪ من النصوص للاختبار. فمثلا لو أردنا تصنيف مدونة الخطابات الإدارية فعلينا أن نختار ٣٦٠ خطاب من مجموع الخطابات الـ ٤٠٠ وإبقاء ٤٠ خطاب لاختبار مدى كفاءة الآلة. ويتم تغييب بيانات الاختبار عن الآلة، وعند اختبار الآلة فإن هذه البيانات تظهر بدون تصنيف مسبق، وتقوم الآلة بمحاولة تصنيفها باستخدام ما تعلمته من المشاهدات السابقة في بيانات التدريب.

وعليه، قمتُ بتقسيم المدونات الثلاث إلى ١٠٪ للاختبار و٩٠٪ للتعلم، وقمنا بتدريب الآلة باستخدام خوارزميات مختلفة عن الأولى؛ وهي: خوارزمية Naïve Bayes، وقمنا باستخدامها على أنها خوارزمية خطية. والخوارزمية الأخرى هي خوارزمية شعاع الدعم الآلي Support Vector Machines - SVM حيث تتمكن هذه الخوارزمية من إيجاد أُطر غير خطية لتصنيف الوثائق.

وأظهرت نتائج تدريب الآلة باستخدام الخوارزميات بأن تصنيف وثائق المدونات أمر ممكن، وأظهرت أيضا بأن بيانات التعلم هي بيانات ذو جودة عالية حيث يسهل على الآلة التفريق بين وثائق المدونات الثلاث، ونرى هذه النتائج في الجدول (١٢)، ويوضح هذا الجدول بأن الآلة قد دُرّبت على استخدام ١٠٨٠ مقال وذلك باختيار ٣٦٠ مقال بشكل عشوائي من كل مدونة، كما يوضح الجدول بأن الآلة قد قامت بتصنيف ١٠٢٣ مقال بشكل صحيح، وذلك باستخدام خوارزمية Naïve Bayes، ويعني ذلك أن الخوارزمية قد فشلت بتصنيف ٥٪ من المقالات. و

ولذلك يحكم على كفاءة الخوارزمية بأنها ٩٥٪ بينما نجحت خوارزمية SVM بتحقيق نتيجة أعلى وذلك بكفاءة تقارب ٩٩٪ مما يدل على أن خوارزمية SVM تُوظّف تقنية تتيح لها التفريق بين وثائق المدونات الثلاث بجودة أعلى. وفي الحالتين: قمنا بتدريب الآلة على استخراج مميزات features تتيح لها تحديد نوع الوثيقة، وعليه؛ بما أنني تعاملت مع نصوص كانت المميزات فيها متعلقة باللغة الطبيعية، أفادت كلتا الخوارزميات باستخلاص كلمات وحروف تتيح لها جميعا التمييز بين الوثائق، وذلك عن طريق استخدام تردد المصطلح معكوس التردد للوثيقة TF-IDF، وهو معامل إحصائي يستخدم لقياس أهمية وجود كلمة في وثيقة ما داخل مجموعة كبيرة من نصوص المدونة.



وتزداد أهمية الكلمة بزيادة عدد مرات ظهور الكلمة أو المصطلح في نص الوثيقة، غير أن تجنب إعطاء أهمية كبيرة للكلمات الشائعة كأحرف الجر يدفع معامل حساب تكرار الكلمة في المدونة بأكملها إلى اعتبار الكلمات التي تظهر في جميع الوثائق بأنها كلمات شائعة ليست ذات أهمية. ويساعد معامل TF-IDF الخوارزمية في استخراج الكلمات (المميزة) ذات الأهمية العالية لأجل أن تتيح للآلة مقدرة التفريق بين الوثائق.

ولاختبار مدى جودة الخوارزميات يجب اختبارها باستخدام وثائق لم ترها الآلة من قبل ولم تستخدم في عملية التدريب، وتعرف هذه الوثائق باسم «بيانات الاختبار». وكما أسلفنا سابقًا، قمتُ باختيار ٤٠ مقالاً من كل مدونة للاختبار، ومثل هذا الاختيار ١٠٪ من مقالات كل مدونة. كما قمتُ باختبار الآلة التي تم تدريبها مسبقا باستخدام خوارزميات Naïve Bayes و SVM ؛ لنختبر مدى قدرة الآلة على التنبؤ بهذه الوثائق وتصنيفها. ويوضح الجدول (١٣) بأن خوارزمية SVM قامت بتصنيف الوثائق بشكل أفضل من Naïve Bayes، وكان ذلك متوقعًا، كون نتائج التدريب قد أظهرت نفس ذلك. كما يظهر الجدول أيضا تقارب نتائج التدريب والاختبار لكل خوارزمية ليدل على أن عملية التدريب قد تمت بشكل صحيح.

الخوارزمية	Naïve Bayes	SVM
عدد وثائق بيانات التعلم	١٠٨٠	١٠٨٠
عدد المدونات	٣	١٤
عدد الوثائق في كل مدونة	٣٦٠	٣٦٠
حالات صنفت بشكل صحيح	١٠٢٣	١٠٦٦
حالات صنفت بشكل خاطئ	٥٧	١٤
كفاءة الخوارزمية	٪٩٥	٪٩٨,٧

الجدول (١٢): نتائج التدريب على تصنيف المدونات باستخدام التعلم الآلي.



الخوارزمية	Naïve Bayes	SVM
عدد وثائق بيانات الاختبار	١٢٠	١٢٠
عدد المدونات	٣	٣
عدد وثائق الاختبار لكل مدونة	٤٠	٤٠
حالات صنفت بشكل صحيح	١١٥	١١٧
حالات صنفت بشكل خاطئ	٥	٣
كفاءة الخوارزمية	٪٩٥,٨	٪٩٧,٥

الجدول (١٣): نتائج الاختبار لتصنيف المدونات باستخدام التعلم الآلي.

وبعد معاينة الأخطاء التي ارتكبتها الآلة خلال تصنيفها للوثائق، وبعد التصدي لمعظم الأخطاء نتيجة خلط الآلة بين مقالات الأخبار الاقتصادية ومقالات أخبار الأسهم، قاربتُ بين نتائج المدونتين. وعليه؛ استطعتُ استخدام لغويات المدونة الحاسوبية لاستخراج مميزات أكثر قد تساعد الآلة على التمييز بين الوثائق. ومن ضمن هذه المميزات استخدام معامل الاحتمال اللوغاريتمي Log Likelihood وهو شبيه بـ TF-IDF الذي يحاول حساب أهمية كلمة معينة بالرجوع إلى نص مماثل للمقارنة. ومن المميزات الأخرى التي يمكن استخدامها أيضا هو نسبة نوع الرمز المميز Type-Token Ratio – TTR، وهو ما يستخدم عادة لتحديد هوية الكاتب. والمثال على ذلك: معرفة جنس الكاتب سواء كان ذكرًا أو أنثى (Homes 1994). ويتم حساب TTR عن طريق قسمة العدد الكلي للكلمات المختلفة (الأنواع – types) على العدد الكلي للكلمات الفعلية tokens. وتدل قيمة TTR على درجة التباين المعجمي، إذ كلما زادت القيمة دل ذلك على وجود درجة عالية من التباين، وهو ما قد يساعد الآلة على التفريق بين وثائق المدونات.

وقد يطول الحديث حول المميزات features التي يمكن استخراجها لمساعدة الآلة؛ وذلك لكثرتها في لغويات المدونات الحاسوبية، ويعتمد اختيار هذه المميزات على



التجربة؛ ولذلك فإن على الباحث أن يجري بعض التجارب باستخدام مميزات مختلفة ومن ثم تصنيف هذه المميزات وترتيبها حسب الأهمية، ويظهر ذلك عن طريق حساب كفاءة الآلة باستخدام كل ميزة على حده، ومن ثم حساب الكفاءة بضم المميزات في مجموعات صغيرة إلى أن يصل الباحث إلى نتيجة ترفع من أداء الآلة.

٢. ١٠. الخاتمة

هناك تطبيقات حاسوبية تجريبية كثيرة تم ذكرها في هذا الفصل بشكل عام. وألقى هذا الفصل الضوء على طريقة بناء المدونات الصغيرة، والمعروفة بالمدونات المختصة، وأوضحتُ ثلاث طرق هي على النحو الآتي: الويب كمدونة، والمدونات أخرى، وتقشير شبكة المعلومات. وكان التركيز الأكبر في هذا الفصل على اللغة العربية الإدارية حيث ارتبطت المدونات الثلاث التي تم جمعها بالعربية الإدارية وكيفية خطاب الإدارة المُتشكّل برئيس مجلس الإدارة والرئيس التنفيذي للمساهمين من ذكور وإناث. وقارنتُ في هذا الفصل بين العربية الإدارية المستخدمة في المدونات الثلاث عن طريق استخدام أدوات تحليل المدونة الحاسوبية، واستخدمتُ أكثر من أداة مجانية تساعد القارئ على إعادة هذه التحليلات باستخدام مدوناتهم الخاصة. وأظهرت التحليلات المبدئية في هذا الفصل عن وجود خلل وفجوة في تواصل وخطاب الإدارة مع المساهمين، وكيف أن المساهمين يتجهون إلى مصادر أخرى كالأخبار الاقتصادية وأخبار الأسهم؛ من أجل معرفة الوضع المالي للشركة. وخلص الفصل إلى إجراء تجارب على تصنيف وثائق المدونات آليا وذلك باستخدام التعلم الآلي، حيث أظهرت النتائج أن باستطاعة الآلة- ببيانات التدريب الجدية-نسخَ التصنيف اليدوي لخبير في المدونات. ويوضح الفصل أهمية لغويات المدونة الحاسوبية في دراسة لغة ما حيث توفر الأدوات والبرامج المعدة مسبقا كثيرًا من الوقت والجهد على المُحلّل.



قائمة المراجع

المراجع الأجنبية

- Al-Thubaity, A., & Al-Mazrua, M. (2014a). Khawas: Arabic Corpora Processing Tool USER GUIDE. Retrieved July 2, 2016, from http://www.sourceforge.net/projects/kacst-acptool/files/?source=navbar.
- Al-Thubaity, A., Khan, M., Al-Mazrua, M., & Almoussa, M. (2014b). KACST Arabic Corpora Processing Tool ‘‘Khawas’’ [Computer Software]. Retrieved July 2, 2016, from https://sourceforge.net/projects/ghawwasv4/.
- Anthony, L. (2014). AntConc (Version 3.4.3) [Computer Software]. Tokyo, Japan: Waseda University. Available from http://www.laurenceanthony.net/
- Björnsson, C. H. (1968). Läsbarhet. In Stockholm: Liber.
- El-Haj, M., and Rayson (2016). «OSMAN - A Novel Arabic Readability Metric». 10th edition of the Language Resources and Evaluation Conference (LREC›16). May 2016. Portoroz, Slovenia.
- El-Haj, M., Koulali, R. (2013). «KALIMAT a Multipurpose Arabic Corpus» at the Second Workshop on Arabic Corpus Linguistics (WACL-2) 2013.
- El-Haj, M., Rayson, P., Young, S., Moore A., Walker M., Schleicher T. and Athanasakou V. (2016). «Learning Tone and Attribution for Financial Text Mining». 10th edition of the Language Resources and Evaluation Conference (LREC›16). May 2016. Portoroz, Slovenia.
- Gunning, R. (1968). The Technique of Clear Writing. McGrawHill.
- Holmes, D. (1994). Authorship Attribution. Computers and the Humanities, 28(2):87–106.
- Kilgarriff, A. (2001). Web as corpus. Reprinted in: Sampson, G. and McCarthy, D. (eds.) (2004) Corpus linguistics: readings in a widening discipline. London: Continuum, 471–473.
- Kincaid, J., Fishburne, R., Rogers, R., and Chissom, B. (1975).



Derivation of new readability formulas (automated readability index fog count and flesch reading ease formula) for navy enlisted personnel. In Research Branch Rep, pages 8–75, Memphis, TN.

- McEnery, T. & Wilson, A. (1996). Corpus linguistics. Edinburgh University Press

- Partington, A. (2003) The Linguistics of Political Argumentation: The Spin-doctor and the Wolf-pack at the White House. London: Routledge.

- Partington, A. (2004) 'Corpora and Discourse, a Most Congruous Beast', in A. Partington, J. Morley and L. Haarman (eds) Corpora and Discourse, pp. 11–20. Bern: Peter Lang.

- Sebastiani, F. (2002). Machine learning in automated text categorization. ACM Comput. Surv.34, 1 (March 2002), 1-47.


هذه الطبعة
إهداء من المركز
ولا يسمح بنشرها ورقياً
أو تداولها تجارياً

الفصل الثالث

د. عبدالمحسن بن عبيد الثبيتي

المعالجة الآلية للكلمات المميزة للمدونات: قضايا تقنية

الملخص

تُعرّف الكلمات المميزة ضمن إطار لسانيات المدونات اللغوية بالكلمات التي يختلف تكرارها في المدونة اختلافا واضحًا عن تكرارها في مدونة أخرى وذلك من منظار إحصائي. ولاستخراج الكلمات المميزة للمدونة فوائد واستخدامات متعددة تبدأ من توجيه اهتمام الباحث لما يمكن دراسته وليس انتهاء بتحديد الأخطاء والفروق الإملائية بين المدونات. ويتطرق هذه الفصل بشكل أساسي إلى التعريف أولا بالكلمات المميزة واستخداماتها وفوائدها وكيفية استخراجها من المدونة من جهة. كما يتطرق إلى تأثير المدونة المرجعية reference corpus من حيث حجمها وموضوعاتها وزمنها ومنطقتها الجغرافية على نتائج الكلمات المميزة من جهة ثانية. ويؤصل لأهم المقاييس الإحصائية المستخدمة في استخراج الكلمات المميزة مع تبيان تأثيرها على الكلمات المميزة وأوجه الشبه والاختلاف بين هذه المقاييس من جهة ثالثة. ومن جهة رابعة: يشرح بشكل مبسط ما يمكن استبعاده من الكلمات المميزة اعتمادا على نوع الدراسة المستخدمة وكمية الكلمات المميزة التي يمكن اعتبارها في الدراسة. وقد درستُ التأثيرات المذكورة آنفا باستخدام منهج تجريبي، مُركزا على الجانب الكمي من موضوع الكلمات المميزة، ولم أتطرق إلا بشكل بسيط جدا إلى تحليل الكلمات المميزة المستخرجة أثناء التجارب



لتشعب هذا الموضوع وارتباطه بنوع الدراسة المعتمدة على الكلمات المميزة. وأهم ما خلصت إليه التجارب هو أن التأثير الأكبر للمدونة المرجعية يتعلق بموضوعاتها ومدى تنوعها بينما حجمها وزمنها ومنطقتها الجغرافية قليلة التأثير. كما أوضحت التجارب أن المقاييس الإحصائية يمكن أن تصنف إلى مجموعتين، كل مجموعة منها تضم مقاييس تتشابه بشكل كبير جدا في الكلمات المميزة الأكثر تميزا، وتكشف عن نمط معين من الكلمات المميزة. وتتضمن المجموعة الأولى معامل الاحتمال اللوغاريثمي، ومربع كاي، ومعامل زد، ومعامل تي. بينما تتضمن المجموعة الثانية معامل الغرابة والمعلومات المتبادلة. وفي نهاية الفصل ذكرت بعض الملاحظات والاعتبارات التي يجب النظر إليها بعد استخراج الكلمات المميزة.

الكلمات المفتاحية: الكلمات المميزة، مدونة الدراسة، المدونة المرجعية، لسانيات المدونات، المقاييس الإحصائية، القيمة الحرجة.

٣. ١. المقدمة

للسانيات المدونات اللغوية (المدونات فيما بعد) شقان، يعتمد ثانيهما على أولهما، ولا قيمة لأولهما بلا ثانيهما. فالشق الأول كمي إحصائي، أما الثاني فتحليل كيفي لبيانات الشق الأول ضمن إطار نظري معين يجعل من هذه البيانات مؤثرة ضمن مجالها، وقد يتعداه إلى غيرها. ويعتمد الشق الكمي بشكل أساس على تكرار الكلمات أو التراكيب في المدونة [1]، مثل النظر في الكلمات الشائعة في المدونة أو الكشاف السياقي لهذه الكلمات، وعلى مقارنة التوزيع الإحصائي لهذه الكلمات بين مدونتين أو بين الكشاف السياقي لهذه الكلمات وبقية المدونة التي ظهرت فيها هذه السياقات. وبعبارة أخرى، فإن أي دراسة ضمن لسانيات المدونات يجب أن تقوم على شق كمي يتضمن كل أو بعض البيانات، وهي على النحو الآتي: تكرار كلمات المدونة، وتكرار الكلمات داخل السياق، وموضعها داخل السياق، والكلمات المميزة للمدونة التي تعد محل الدراسة مقارنة بمدونة أخرى، والتصاحب أو التلازم.

والجدير بالذكر هنا هو أن النتائج الكمية نفسها يمكن أن تحلل كيفيا وتدرس بواسطة نظريات لسانية مختلفة. فعلى سبيل المثال، فإن بيانات التكرار الخاصة بكلمات

١ – سأكتفي لاحقا بالكلمات فقط دون ذكر التراكيب؛ لأن ما يمكن أن يتم مع الكلمات يمكن أن يتم مع التراكيب.



المدونة-خاصة الكلمات الأكثر تكرارا- يمكن أن تستخدم في تصميم المناهج الدراسية سواء لأهل اللغة أو لمتعلميها كلغة ثانية ويمكن أن تستخدم أيضا في بناء مداخل المعاجم بحسب الغرض منها، وتكون بيانات الكشاف سياقي مساعدًا على وضع الأمثلة المناسبة لهذه المداخل. ويُمكن أيضا لبيانات التصاحب اللفظي أن تكون ضمن شواهد بعض المعاجم، كما يمكن أن تستخدم أيضا في نظريتي التهيئة المعجمية (ماكنري وهاردي ٢٠١٦: ٢٧١-٢٧٤) والنظم الدلالي (المصدر السابق: ٢٥٤-٢٦٥).

ومن أجل أن يكون التحليل الكيفي ذا قيمة يتمكن بها الدارس من الوصول إلى الاستنتاج المناسب وفهم الاختلاف في النتائج عند استخدام أكثر من طريقة إحصائية لنفس الغرض، فإنه يلزمه الإلمام بالنواحي الأساسية للطرق الكمية والإحصائية المستخدمة في الشق الكمي من لسانيات المدونات (التكرار، والكشاف السياقي، والكلمات المميزة، والتصاحب أو التلازم) الذي يبني عليه تحليله الكيفي. وسوف يتطرق هذا الفصل إلى جانب من جوانب الشق الكمي في لسانيات المدونات ألا وهو استخراج الكلمات المميزة للمدونة، أي تحديد الكلمات التي تميز المدونة الخاضعة للدراسة عند مقارنتها بمدونة أخرى.

وعلى الرغم من أن الكلمات الأكثر تكرارا تشبه البصمة في تحديد هوية المدونة، إلا أن الكلمات الأكثر تكرارا في المدونات قد تتشابه بدرجة كبيرة. ففي المدونات الصحفية-على سبيل المثال-قد نجد الكلمات الوظيفية، أو بعض الكلمات مثل صرح، وقال، ووزير، ولفظ الجلالة (الله)، وأسماء القادة السياسيين ضمن الكلمات الأكثر تكرارا. كما أن المدونات التي تتشابه في الموضوع قد تتطابق تطابقا كبيرا في كلماتها الأكثر تكرار، وعلى وجه الخصوص: إذا جمعت بنفس مقاييس التصميم. ولو ألفينا ضمن الكلمات الأكثر تكرارا في مدونة ما كلماتٍ لم ترد ضمن الكلمات الأكثر تكرارا في مدونات أخرى، فهل يعني هذا أن لظهورها مغزى ضمن إطار دراستنا الحالية؟ أم أنه اختيار لفظي فقط يمكن أن تقوم به كلمة أخرى؟ خذ على سبيل المثال: كلمتي (بنوك) و(أبناك) اللتين تدلان على نفس المعنى ولكن الأولى مستخدمة في الصحافة الاقتصادية السعودية، والثانية في الصحافة الاقتصادية المغربية. ومثلهما كلمتا (مظاهرة) و(تظاهرة) والمتتابعات اللفظية (رئيس الوزراء) و(الوزير الأول) وهلم جرا. ولا تخلو المدونات من أنماط خفية غير ظاهرة في الكلمات الأكثر تكرارا، وعليه؛ كيف يمكننا اكتشافها؟ لو



افترضنا أن الباحث كان مهتما بالتحليل النقدي للخطاب مثلا، واستطاع أن يستعرض قائمة التكرار للمدونة التي يدرسها، فما هو المنهج الذي سوف يختاره لتحديد الكلمات التي سوف ينطلق منها لدراسته؟ وهل سيكون اختياره بعيدا عن الانحياز لفكرة أو أيدلوجية معينة؟

وللتخلص من الإشكالات المثارة سابقا، يمكننا استخدام ما يسمى بالكلمات المميزة للمدونة، وهي الكلمات التي يختلف تكرارها في مدونة الدراسة primary corpus من وجهة نظر إحصائية عن مدونة أخرى (المدونة المرجعية reference corpus) (سكوت وتريبيل Scott and Tribble 2006). لكن، ما الذي سيُفيد الباحث من هذه المعالجة؟ إن لمعرفة الكلمات المميزة للمدونة العديد من الفوائد والتطبيقات[1]، ومن أهمها ما أسوقه على النحو الآتي:

التعرف على السمات العامة للمدونة التي ندرسها. هل هي مدونة متخصصة في حقل معرفي معين (اللغة، الفيزياء، الاقتصاد، إلخ.)، أو موضوع صحفي معين (اقتصادي، سياسي، رياضي، إلخ.)، أو أن لها تفضيلات لغوية معينة، وهلم جرّا (كيلغاريف Kilgarriff 2012).

معرفة الفروق اللغوية في الدراسات المقارنة بين المدونات مثل المقارنة بين الصحافة المغربية والصحافة السعودية، أو بين اللغة المستخدمة في الأبحاث الطبية والنصائح والإرشادات التي تقدم للمرضى من قبل الجهات الطبية (الشمري والثبيتي 2015، وغرابوفسكي Grabowski 2013).

حصر وتحديد الكلمات التي يجب التركيز عليها في الدراسة؛ لأن هذه الكلمات هي منطلق الدراسة اللغوية، وخصوصا في مجال التحليل النقدي للخطاب critical discourse analysis (الشمري والمحمود «الفصل السادس»، وبيكر Baker 2010).

البعد عن التحيز لأفكار مسبقة أو التعامل مع القضايا اللغوية بناء على أيديولوجيات معينة، كون تحديد ما يستوجب الدراسة يعتمد اعتمادا تاما على المقاييس الإحصائية وما يعكسه تكرار الكلمات داخل المدونة دون أي اعتداد بالاختيارات الشخصية المبنية على

١– من وجهة نظري فإن الاستفادة من منهج لسانيات المدونات ونتائج الأساليب الكمية المستخدمة فيها لا تُضفي قيمةً بحثية إلا بمدى سعة مدارك الباحث، شريطة التزام هذا الباحث بالمنهج العلمي.



الحدس والتخمين والأفكار المسبقة (بيكر Baker 2004، والشمري والثبيتي 2015).

هناك استخدامات أخرى مثل التعرف على الأخطاء الإملائية الشائعة في مدونات متعلمي اللغة مثلا عند مقارنتها بمدونة تم اختيار نصوصها بعناية، وكذلك مثل معرفة ما يميز كتابات متعلمي اللغة من جنسية معينة مقارنة بكتابات المتحدثين الأصليين للغة ذاتها أو لمتحدثي لغات أخرى (غرانجر وباكوت Granger and Paquot 2013)، أو معرفة الأخطاء في نصوص المدونة نتيجة نقل البيانات أو تغيير ترميزها عند مقارنتها بالمدونة نفسها قبل النقل أو تغيير الترميز، أو معرفة الكلمات أو الوسوم التي ليست من النصوص عند جمع المدونة آليا عن طريق الإنترنت ليتم بعد ذلك إزالتها من نصوص المدونة (كيلغاريف Kilgarriff 2012).

ويتطلب استخراج الكلمات المميزة-كما يبين التعريف المذكور سابقا-مدونةً مرجعية ومقياسًا إحصائيًا لحساب الاختلاف في تكرار الكلمات بين المدونتين والتأكد من أنه اختلاف معتبر إحصائيا. وفيما يلي الخطوات المتبعة لاستخراج الكلمات المميزة:

الأولى: قائمة بالكلمات وتكرارها في مدونة الدراسة (المدونة الرئيسة).

الثانية: قائمة بالكلمات وتكرارها في المدونة المرجعية.

الثالثة: حسابات إحصائية للكشف عن الاختلافات في التوزيع الإحصائي لتكرار الكلمات بين مدونة الدراسة والمدونة المرجعية. وبعد أن يتم تطبيق المقياس الإحصائي يتم ترتيب الكلمات تنازليا حسب قيم المقياس الإحصائي، ثم تتم تصفية النتائج بحسب الخطوات التالية:

الأولى: يتم اختبار الكلمات التي تكون قيم المقياس الإحصائي لها أكبر أو مساوية لقيمة معينة تسمى القيمة الحرجة critical value. وتحديد هذه القيمة مرتبط بما يسميه علم الإحصاء نسبة الخطأ (أو) درجة الموثوقية، وأقل نسبة خطأ يمكن قبولها هي ٠,٠٥ وتقابلها القيمة ٠,٩٥ لدرجة الموثوقية. ومقابل كل نسبة خطأ أو درجة موثوقية قيمة حرجة لكل مقياس إحصائي بحيث تقبل الكلمة التي تحقق هذه القيمة أو أعلى منها. وتتوفر هذه المعلومات عادة في الجداول الإحصائية الخاصة بكل مقياس إحصائي.

الثانية: يُعلي بعض المقاييس الإحصائية من قيمة الاختلاف في التوزيع الإحصائي للكلمات التي يكون تكرارها كبيرًا جدًا في مدونة الدراسة مقارنة بالمدونة المرجعية، أو



للكلمات التي يكون تكرارها ضئيلاً جدًا في مدونة الدراسة مقارنة بالمدونة المرجعية. ولذلك يجب التأكد بعد تطبيق نسبة الخطأ (أو) درجة الموثوقية من أن تكون نسبة تكرار الكلمة المميزة في المدونة الرئيسية مقارنة بنسبة تكرارها في المدونة المرجعية أكبر من حد معين.

الثالثة: تتضمن قائمة الكلمات المميزة بعد تطبيق الخطوتين السابقتين كلمات لم تظهر سوى مرة أو مرتين في المدونة الرئيسة (مدونة الدراسة) غير أنها لم تظهر أبدا في المدونة المرجعية. ومثل هذه الكلمات لا يمكن الاعتداد بها ككلمات مميزة، لذا يتم تصفية النتائج بحذف الكلمات التي يقل تكرارها عن حد معين.

وليس لاختيار نسبة الخطأ ونسبة تكرار الكلمة المميزة في المدونة الرئيسية مقارنة بنسبة تكرارها في المدونة المرجعية ضابط صارم معلوم، بل يخضع لاجتهاد الباحث وطبيعة البحث والمدونات التي يستخدمها. ولكن الباحث ملزم بتوضيح الأسباب التي دعته لتلك الاختيارات.

وسأستعرض في المبحث (٣. ٣.) تأثير المدونة المرجعية من حيث حجمها، ونوع الموضوعات التي تغطيها، والفترة الزمنية التي جمعت فيها، والمنطقة الجغرافية التي صدرت منها، وذلك كله على الكلمات المميزة. أما في المبحث (٣. ٤.) فسأعرّج على تأثير أشهر المقاييس الإحصائية المستخدمة في الكشف عن الكلمات المميزة والاختلافات فيما بينها. وسيتضمن المبحث (٣. ٥.) الخطوات التي يمكن القيام بها بعد استخراج الكلمات المميزة. أما في المبحث (٣. ٢.) من هذا الفصل، فسأوضح المدونات التي سأستخدمها في التجارب، مع شرح طريقة معالجتها قبل استخدامها، وشرح كيفية حساب قيم المقاييس الإحصائية بالاعتماد على ما يسمى بجدول المصادفات.

وأخيرًا أنوه إلى أن الهدف من هذا الفصل هو معرفة الجانب الكمي من قضية الكلمات المميزة وما يتعلق به فقط دون الدخول في أي تحليل لغوي. فكما أشرت سابقا فإن نفس النتائج يمكن تحليلها والنظر إليها من وجهات نظر الاتجاهات اللغوية المختلفة.



٣. ٢. البيانات والأدوات

٣. ٢. ١. المدونات

سأستخدم في هذه الدراسة المدونتين الآتيَتَيْن:

أولا: مدونة الصحافة العربية ٢٠١٥؛ وتشمل ما ورد في صحف سبع دول عربية هي السعودية، ومصر، واليمن، والسودان، وتونس، والجزائر، والمغرب خلال الفترة من اليوم الأول لعاصفة الحزم إلى اليوم الأخير منها. وتم جمع هذه النصوص آليا من مواقع الويب: سعورس(١)، ومصرس(٢)، ويمرس(٣)، وسودارس(٤)، وتورس(٥)، وجزايرس(٦)، ومغرس(٧). ويُوضح الجدولان (١) و(٢) البيانات الإحصائية الأساسية للمدونة بحسب المواضيع والدول على التوالي.

ثانيًا: مدونة الصحف السعودية؛ وهي مدونة جُمعت يدويا من الصحافة السعودية، وتغطي عامي: ٢٠٠٧م، و٢٠١٢م. أما الجدولان (٣) و(٤) فيوضحان البيانات الإحصائية الأساسية للمدونة بين عامي ٢٠٠٧م و٢٠١٢م على التوالي.

المواضيع	عدد النصوص	الحجم	عدد الكلمات
الاجتماعية	١١,٤٢٧	٢,٩٦٠,٩٩٨	٧٤,٣٨٧
الاقتصادية	٩,٢٥٠	٣,٠٦٨,١٣٨	٦٧,٦٥٢
الثقافية	٧,٧٠٥	٢,٨٩٩,٧٤٤	٩٤,٥٤٣

١ - انظر: https://www.sauress.com.

٢ - انظر: https://www.masress.com.

٣- انظر: https://www.yemeress.com.

٤ - انظر: https://www.sudaress.com.

٥ - انظر: https://www.turess.com.

٦ - انظر: https://www.djazairess.com.

٧- انظر: https://www.maghress.com.



المواضيع	عدد النصوص	الحجم	عدد الكلمات
الدولية	١٣,٢٥٢	٤,٣٠١,٧٤٧	٨١,٠٤٩
الدينية	٢,٥٥٩	١,٤٩١,٩٤٦	٥٧,٢٩٠
الرياضية	١٣,٠٤٨	٣,٢٥١,٤٧١	٧٣,٢٧٣
السياسية	١٤,٦٤٧	٥,٣٧٩,٤٩٩	١٠١,٦٨٠
الصحية	٤,٥٢٢	١,٤٢٣,٣٦٣	٤٦,٧٨٥
المجموع	٧٦,٤١٠	٢٤,٧٧٦,٩٠٦	٢٣٩,١٥٣

الجدول (١): توزيع كلمات مدونة الصحف العربية على الموضوعات.

الدولة	عدد النصوص	الحجم	عدد الكلمات
السعودية	١٣,٣٢٤	٤,٢٣٦,٧٥٥	٨٤,٧٧٥
مصر	٢٩,٧٩٨	٨,٦١٠,٤٨٣	١٢٩,٣٤٧
اليمن	٥,٢٠٩	١,٨٢٤,٤٨٤	٦٠,١١١
السودان	٥,٠٣٤	٢,٤٩٤,٣٤٥	٨٥,٢٥٢
تونس	٤,٦٠٨	١,١٤٧,٩٩٤	٤٧,٥٦٩
الجزائر	٨,٢٦٩	٢,٨٣٢,٥٠٥	٧٦,٩٣٦
المغرب	١٠,١٦٨	٣,٦٣٠,٣٤٠	٨٩,٥٥٧
المجموع	٧٦,٤١٠	٢٤,٧٧٦,٩٠٦	٢٣٩,١٥٣

الجدول (٢): توزيع كلمات مدونة الصحف العربية على الدول.



الموضوع	عدد النصوص	الحجم	عدد الكلمات
الاجتماعية	٧٣١	٣١٦,٤٣٦	٢٨,٥٠٢
الاقتصادية	٧٣٩	٣٧٩,٤٠٦	٢١,٣٨١
الثقافية	٧٣٨	٥٣٣,٥٢٣	٤٠,٢٣٠
عام	٧٢٨	٥٢٠,١٢٥	٤٠,٢٤٢
الرياضية	٧٣١	٢١٥,٨٤٢	١٨,٣٢٦
السياسية	٧٢٦	٤١٢,٠٥٤	٢٦,٢٠٣
التقنية	٧٢٨	٤٩٣,٨٩١	٣١,٨٠٦
المجموع	٥,١٢١	٢,٨٧١,٢٧٧	٨٧,٣٦٤

الجدول (٣): توزيع كلمات مدونة الصحف السعودية ٢٠٠٧م.

الموضوع	عدد النصوص	الحجم	عدد الكلمات
الاجتماعية	٧٠٠	٢١٩,٢١٩	٢١,٦٢٨
الاقتصادية	٧٠٠	٣٤١,٢٠٠	٢١,٥٤٢
الثقافية	٧٠٠	٣٣٣,٢٤٥	٣٠,١٧٦
عام	٦٩٨	١٥٢,٥٥٧	١٩,٦٤٠
الرياضية	٧٠٠	٢٢٤,٨٨٩	١٧,٧٣٧
السياسية	٧٠٠	٢٥٢,٤٠٣	١٩,٧٤٥
التقنية	٧١٥	٣٠٢,١٩٣	٢٢,٤٢٩
المجموع	٤,٩١٣	١,٨٢٥,٧٠٦	٦٧,٣٨٥

الجدول (٤): توزيع كلمات مدونة الصحف السعودية ٢٠١٢م.



وقد استخدمت مُقطّع ستانفورد[١] (مونرو وجرين ومانينج Monroe, Green and Manning 2014) الذي يقوم بفصل المورفيمات المحيطة بالكلمة -إن وجدت- مثل الباء والكاف والفاء والضمائر المتصلة-على سبيل المثال-لتقليل تشتت الكلمات. فكلمات مثل (كتاب، وكتاب، فكتاب، ولكتاب، وكتابهما، وكتابه) سوف تفصل عنها الواو، والفاء، والواو والام، والواو وهما، والهاء، وبالتالي سوف يزداد تكرار كلمة كتاب في المدونة ويظهر تكرارها الفعلي. وعلى الرغم من أن التقطيع يزيد من حجم المدونة (العدد الكلي للكلمات) بسبب المورفيمات التي تم فصلها (جميعها من الكلمات الوظيفية) إلا أنه يقلل من عدد الكلمات غير المكررة في المدونة بنسبة ٦٠٪ تقريبا (الثبيتي والسبيعي Al-Thubaity & Al-Subaie 2015). ويقوم البرنامج أيضا بإزالة التشكيل وتحويل الألف المهموزة إلى ألف غير مهموزة، ولكن النتائج التي نجنيها من البرنامج تفوق الأضرار الناتجة عن مثل هذه التغييرات، وخصوصا في كونها غير مؤثرة على فهم السياقات التي ترد فيها الكلمات، إضافة إلى أن نوعية الأخطاء التي ترد في البرنامج لها نمط ثابت ومحدود.

٣. ٢. ٢. المقاييس الإحصائية

هناك الكثير من المقاييس الإحصائية التي يمكن استخدامها لاستخراج الكلمات المميزة. وأشهر هذه المقاييس وأكثرها استخداما في لسانيات المدونات: الاحتمال اللوغاريثمي ومربع كاي، ويعود السبب في هذا إلى أن العديد من البرامج التي صممت لمعالجة المدونات وفرتها هي بالذات دون غيرها، وإن كان هذا الحال قد بدأ بالتغير خصوصا مع أنظمة معالجة المدونات، وعلى الأخص ما هو متوفر منها في الشبكة (الويب).

وتعتمد حسابات المقاييس الإحصائية على معلومات التكرار الفعلية token للكلمة في المدونتين الرئيسة والمرجعية، وتسمى هذه التكرارات بالتكرارات المشاهدة، وعلى التكرارات المتوقعة للكلمة والتي يتم حسابها بناء على حجم المدونتين وتكرار الكلمات الأخرى في المدونة. ويتم ترتيب هذه القيم في جدولين هما جدول الاحتمالات للتكرارات المشاهدة وجدول الاحتمالات للتكرارات المتوقعة، ويرمز (ن) إلى حجم

١ - انظر: http://nlp.stanford.edu/software/segmenter.shtml.



المدونة، أما (ش) فيرمز إلى التكرارات المشاهدة أو الملحوظة، أما الرمز (ت) فيشير إلى التكرارات المتوقعة، والرمز (ع) يشير إلى حواصل العلاقات الحسابية بين التكرارات المشاهدة والمتوقعة (انظر الجدولين ٥ و٦). أما الجدول (٧) فيوضح طريقة حساب المقاييس الإحصائية للكلمة لستة من المقاييس الإحصائية التي يمكن استخدامها لاستخراج الكلمات المميزة. وبواقع المعالجة الإحصائية؛ هناك العديد من المقاييس التي يمكن استخدامها ودراسة صلاحيتها لاستخراج الكلمات المميزة غير أن المذكورة في الجدول (٧) تُعدّ الأشهر، كما أن نطاق هذا الفصل تعريفي وليس مجاله ذكر كل ما يتعلق بمثل هذه المقاييس.

وللحكم على موثوقية نتائج المقاييس الإحصائية الموضحة في الجدول (٧) نستخدم ما يسمى بمستوى الثقة أو نسبة الخطأ وهما مقداران مترابطان جدا. فنسبة الخطأ يجب أن تكون أقل من مقدار معين (أكبر نسبة خطأ يمكن قبولها هي ٠٫٠٥؛ قارن مع المجيول ٢٠١٥: ٢٤٩-٢٥٢)، ويقابلها مستوى ثقة معين (أقل نسبة ثقة يمكن قبولها هي ٩٥٪)، ولكل معيار من هذه المقاييس قيمة حرجة يجب أن تتساوى معها نتائج المعيار الإحصائي أو تتعداها حتى تحقق مستوى الثقة ونسبة الخطأ المطلوبتين. وبطبيعة الحال يمكن الإشارة إلى مستوى الثقة أو نسبة الخطأ فقط بدلا من ذكر الاثنتين. والجدير بالذكر هنا هو أن مقياسيْ معامل الغرابة والمعلومات المتبادلة ليس لهما جداول إحصائية تمكننا من استخراج القيم الحرجة لهما (الجدول ٨ يوضح القيم الحرجة للمقاييس الإحصائية من مستويات الثقة).

وقد استخدمت في جميع التجارب نظام معالجة المدونات العربية غواص الإصدار ٤٫٦ [١] وهو النسخة الأحدث من النظام (الثبيتي وآخرون Al-Thubaity et. al. 2013) ، ويتضمن المقاييس الإحصائية المذكورة آنفا. وقد أستخدمتُ برنامج «المختار العربي» الذي يقوم بالاختيار العشوائي لنصوص المدونة المرجعية بحسب بعض المعايير التي يحددها المستخدم[٢].

١- انظر: https://sourceforge.net/projects/ghawwasv4/files/.

٢- صمم هذا البرنامج الأستاذ عالي القرني في مدينة الملك عبدالعزيز للعلوم والتقنية.



المدونة	الكلمة	بقية الكلمات	المجموع
المدونة الرئيسية	ش١ تكرار الكلمة في المدونة	ش٢= ن١-ش١ مجموع تكرار بقية الكلمات في المدونة	ن١ حجم المدونة
المدونة المرجعية	ش٣ تكرار الكلمة في المدونة	ش٤= ن٢-ش٣ مجموع تكرار بقية الكلمات في المدونة	ن٢ حجم المدونة
المجموع	ع١= ش١+ش٣	ع٢= ش٢+ش٤	ن= ن١+ن٢

الجدول (٥): جدول الاحتمالات للتكرارات المشاهدة.

المدونة	الكلمة	بقية الكلمات	المجموع
المدونة الرئيسية	ت١ = (ن١×ع١)/ ن	ت٢ = (ن١×ع٢)/ ن	ن١
المدونة المرجعية	ت٣ = (ن٢×ع١)/ ن	ت٤ = (ن٢×ع٢)/ ن	ن٢
المجموع	ع١	ع٢	

الجدول (٦): جدول الاحتمالات للتكرارات المتوقعة.

المقياس الإحصائي	المعادلة الرياضية
معامل الاحتمال اللوغاريثمي (ل.ل)	ل.ل=٢×[(ش١×لو(ش١\ت١) + (ش٢×لو(ش٢\ت٢) + (ش٣×لو(ش٣\ت٣) + (ش٤×لو(ش٤\ت٤)]
مربع كاي (م.ك)	م.ك=[(ش١-ت١)٢\ت١] + [(ش٢-ت٢)٢\ت٢] + [(ش٣-ت٣)٢\ت٣] + [(ش٤-ت٤)٢\ت٤]
معامل زد (م.ز)[١]	م.ز=(ش١-ت١) \(ت١) ½ ١



المقياس الإحصائي	المعادلة الرياضية
معامل تي (م.ت)	م.ت=(ش١-ت١)\(ش١) ½
معامل الغرابة (م.غ)	م.غ= (ش١\ن١)/ (ش٣\ن٢)
المعلومات المتبادلة (م.م)	م.م=لو(ش١\ت١)

الجدول (٧): المعادلات الإحصائية للمقاييس الإحصائية.

المقاييس الإحصائية	نسبة الخطأ ومستوى الثقة				
	.05	.025	.01	.001	.000001
	95 %	97.5 %	99 %	99.99 %	99.9999 %
مربع كاي	3.841	5.024	6.635	10.828	24
التشابه اللوغاريثمي	3.841	5.024	6.635	10.828	24
معامل تي	6.314	12.7062	31.82	318.31	310688.1852
معامل زد	1.645	1.960	2.326	3.090	4.7591

الجدول (٨): جدول القيمة الحرجة للمقاييس الإحصائية مقابل نسبة الخطأ ومستوى الثقة.

٣. ٣. تأثير المدونة المرجعية

يتطرق هذا المبحث إلى تأثير المدونة المرجعية من حيث حجمها ومجالها والفترة الزمنية التي تغطيها والمنطقة الجغرافية التي تم جمع نصوصها منها على عدد ونوعية الكلمات المميزة. وقد تطرقت بعض الدراسات السابقة لتأثير المدونة المرجعية على الكلمات المميزة مثل بربر-سارديناه (Berber-Sardinha 2000) الذي استخدم

١ - عندما يرفع أي عدد للأس ½ فهذا يعني الجذر التربيعي للعدد.



مقياس الاحتمال اللوغاريثمي ودرس تأثير زيادة حجم المدونة المرجعية على عدد الكلمات المميزة ووجد أن عدد الكلمات المميزة يكون ثابتا تقريبا عندما يكون حجم المدونة المرجعية خمسة أضعاف أو أكثر من حجم مدونة الدراسة. وقد درس(قوه Goh 2011) -على سبيل المثال- تأثير حجم المدونة المرجعية المنطوقة والمكتوبة مع زمنها ونوعها والمنطقة الجغرافية على الكلمات المميزة، ووجد بحسب التجارب التي أجراها أن حجم المدونة المرجعية هو الأقل تأثير على عدد الكلمات المميزة من بين العوامل الأخرى بينما موضوعها هو الأكثر تأثيرا.

ولكن ما يميز التجارب في هذا الفصل هو أنها أول دراسة تجريبية عربية خاصة بهذا الموضوع، ويضاف إلى ذلك التركيز على متغير واحد في المدونة المرجعية، وتحييد أي مؤثرات أخرى من قبيل تغيير حجم المدونة المرجعية، أما تحييد تأثير الموضوع والزمن والمنطقة الجغرافية عن طريق عدم تغييرها (نفس الموضوع والزمن والمناطق الجغرافية للمدونات المرجعية) فلا تُوجد على حد علمي دراسة بمثل هذا النوع.

وسأوظف في جميع تجارب هذا المبحث مقياس الاحتمال اللوغاريثمي مع نسبة خطأ أقل من 0.000001 (مستوى ثقة أكبر من 99.9999 % ، انظر الجدول ٨)، شريطة أن تكون نسبة الكلمة في مدونة الدراسة إلى نسبتها في المدونة المرجعية أكبر من أو تساوي ١٠، وأن يكون عدد تكرارها في مدونة الدراسة أكبر من أو يساوي ١٠.

٣. ١. ٣. تأثير الحجم

توصي الأدبيات الخاصة بموضوع استخراج الكلمات المميزة أو التي استَخْدَمتْها بأن يكون حجم المدونة المرجعية أكبر من حجم مدونة الدراسة أو على أقل تقدير أن تكون المدونتان متساويتين في الحجم. والسبب خلف هذه التوصية هو أنه كلما زاد حجم المدونة المرجعية ازدادت فرصة ظهور كلمات مميزة جديدة في المدونة الرئيسة. سأقوم هنا بدراسة تأثير حجم المدونة المرجعية على عدد الكلمات المميزة مع محاولة تحييد المؤثرات الأخرى (الزمن، والموضوع، والمنطقة الجغرافية) حيث سيتم استخراج الكلمات المميزة للرياضة من مدونة الصحف العربية ٢٠١٥م عن طريق مقارنتها ببقية موضوعات المدونة. وسوف أجري خمس تجارب يكون فيها حجم المدونة المرجعية ربعَ حجم مدونة الدراسة (٠,٢٥)، ونصفها (٠,٥٠) ومساويًا لها (١)، وضعفها (٢)،



وأربعة أضعافها (٤) ومن بعدُ كامل المدونة (٦,٦٢).

ويبلغ حجم الجزء الخاص بالرياضة في مدونة الصحف العربية لعام ٢٠١٥م ثلاثة ملايين و٢٥١,٤٧١ كلمة (١٣٪ من حجم المدونة) وبالتالي فإن ربع هذا العدد هو ٨١٢,٨٦٨ سيُختار عشوائيا (بواسطة المختار العربي) بحيث يكون نصيب كل دولة هو سُبْع هذا العدد؛ أي: ١١٦,١٢٤ كلمة (لسبع دول)، وسيُؤخذ من كل موضوع من الموضوعات داخل كل دولة بشكل عشوائي سبعة موضوعات؛ أي ١٦,٥٨٩ كلمة. ويتم العمل بنفس الأسلوب للاختيار العشوائي للكلمات بالنسبة لبقية التجارب (يوضح الجدول ٩ عدد الكلمات المستخرجة في كل تجربة بحسب المحددات الموضحة سابقا).

النسبة بين المدونتين	عدد الكلمات المميزة	الزيادة عن الأصغر	نسبة الكلمات المميزة إلى عدد كلمات المدونة المرجعية
0.25	1,090	-	1.49%
0.50	1,637	547	2.23%
1.00	2,301	664	3.14%
2.00	2,913	612	3.98%
4.00	3,242	329	4.42%
6.62	3,319	77	4.53%

الجدول (٩): عدد الكلمات المميزة مقابل حجم المدونة المرجعية بالنسبة لمدونة الدراسة باستخدام مقياس الاحتمال اللوغاريثمي.

وأول ما نلاحظه بالنسبة لعدد الكلمات المميزة هو أن عددها يزداد بازدياد حجم المدونة المرجعية ولكن الزيادة في الفرق بين عددها يتناقص عندما يتعدى حجم المدونة المرجعية الضعف. وهذا يتوافق مع ما توصل إليه (بربر-سارديناه Berber-Sardinha 2000) بأن عدد الكلمات المتميزة يزداد بازدياد المدونة المرجعية ولكنه يبقى ثابتا عندما



يزيد حجمها عن خمسة أضعاف، فلا يكاد يكون هناك فرق. والسبب في ازدياد عدد الكلمات المميزة هو ازدياد قيمة معامل الاحتمال اللوغاريثمي، وازدياد نسبة الكلمة في مدونة الدراسة مقارنة بنسبتها في المدونة المرجعية مع ازدياد حجم المدونة المرجعية، ويستمر هذا الازدياد حتى لا توجد أي كلمة تحقق الشرط الثالث، وهو أن يكون تكرارها في مدونة الدراسة أكبر أو يساوي ١٠ (يوضح الجدول ١٠ قيم مقياس التشابه اللوغاريثمي بالنسبة للكلمات العشر الأكثر تميزا عندما كان حجم المدونة المرجعية يساوي ربع مدونة الدراسة وكيف تغيرت قيمها بازدياد حجم المدونة).

	حجم المدونة المرجعية					
الكلمة	0.25	0.50	1.00	2.00	4.00	6.62
الفريق	4531.3	8354.2	13550.5	20220.1	26793.3	35776.1
فريق	3568.3	6555.7	10799.6	15791.0	21149.8	27742.6
مباراة	3467.4	6475.1	10685.1	16004.2	21834.8	30488.9
المباراة	3165.7	5805.7	9749.2	14394.0	19869.8	27322.8
كرة	2877.9	5387.2	8878.6	13243.0	17766.9	24587.5
المنتخب	2432.2	4357.9	7395.8	10884.7	14543.8	20183.7
النادي	2306.9	3980.7	7177.6	10080.4	13595.3	19034.7
نادي	2136.0	3973.2	6397.5	9364.9	12704.6	17449.9
القدم	2048.5	4197.7	7163.2	10503.6	13880.6	19498.5
اللاعب	2029.9	3693.5	6132.6	9114.5	12488.3	17689.1

الجدول (١٠): ازدياد قيم معامل التشابه اللوغاريثمي مع ازدياد حجم المدونة المرجعية.

ولكن ما نسبة ما تتشارك به الكلمات المميزة في كل حالة من حالات حجم المدونة



بحسب التجارب التي أجريت كما هو موضح أعلاه؟ تشير البيانات كما يوضح الجدول (١١) إلى أن معدل نسبة الكلمات المميزة المشتركة يصل إلى ٩٢,٤٪، ومن ثم فإنه من المهم ملاحظة أن نسبة ما يظهر من الكلمات المميزة-عندما كان حجم المدونة المرجعية ربعَ مدونة الدراسة-قد بلغت ٩٠٪؛ أي: حوالي ٩٩٠ كلمة.

النسبة بني المدونتين	0.25	0.50	1.00	2.00	4.00	6.62
0.25	100.0%	89.7%	91.0%	91.6%	91.2%	90.8%
0.50		100.0%	91.8%	92.7%	92.3%	89.9%
1.00			100.0%	94.0%	94.0%	92.3%
2.00				100.0%	95.3%	93.9%
4.00					100.0%	96.1%
6.62						100.0%

الجدول (١١): نسبة تشارك الكلمات المميزة بحسب حجم المدونة المرجعية.

ويدعونا هذا العدد الكبير من الكلمات إلى طرح التساؤل الآتي: هل تتشارك هذه الكلمات في ترتيبها أيضا؟ بمعنى آخر: ما مقدار تشاركها في المائتي الرتبة الأولى. والداعي لهذا السؤال هو أن المعتاد في دراسة الكلمات المميزة هو أن يتم اختيار أعلى الكلمات المميزة رتبةً بحسب المقياس الاحصائي بعد تطبيق بقية الشروط بسبب العدد الكبير للكلمات المميزة (سأتحدث عن هذا لاحقا في نهاية هذا الفصل)، فإن وجدنا تشابها كبيرا في الرتبة فهل من الضروري أن يكون حجم المدونة المرجعية أكبر أو مساوٍ لحجم المدونة الرئيسة؟ بصيغة أخرى: هل يمكننا الاعتماد على النتائج عندما تكون المدونة المرجعية أصغر من مدونة الدراسة؟

ويعرض الجدول (١٢) نسبة التشارك في المائتي كلمة الأكثر تميزا بحسب حجم

المدونة المرجعية. وبحسب ما توضحه البيانات فإن نسبة التشارك عالية جدا (كانت ٩٢٪ في أقلّ حالاتها). وهذه النتيجة توضح أن حجم المدونة المرجعية ليس له تأثير يذكر على الكلمات الأكثر تميزا. بل إن ترتيب الكلمات المميزة لا يتأثر كثيرًا، فأغلب الكلمات تحافظ على ترتيبها في القائمة، وإن تغير يكون تغيرًا طفيفًا. ويشير الجدول (١٣) إلى طبيعة ترتيب الكلمات العشرين الأكثر تميزا عندما كان حجم المدونة المرجعية أكبر بمقدار ٦٢ ,٦ من حجم مدونة الدراسة، بالإضافة إلى ترتيبها في الأحجام الأخرى الأصغر منها والمستخدمة للمدونة المرجعية.

النسبة بين المدونتين	0.25	0.50	1.00	2.00	4.00	6.62
0.25	100.0%	92.0%	94.0%	94.0%	94.0%	92.0%
0.50		100.0%	95.5%	96.0%	96.0%	92.5%
1.00			100.0%	97.0%	97.0%	93.5%
2.00				100.0%	97.0%	95.0%
4.00					100.0%	96.5%
6.62						100.0%

الجدول (١٢): نسبة التشارك في المائتي كلمة الأكثر تميزا بحسب حجم المدونة المرجعية.

٣. ٣. ٢. تأثير المنطقة الجغرافية

لتوضيح تأثير المنطقة الجغرافية، نستخرج الكلمات المميزة لموضوعات الصحافة السعودية (من مدونة الصحافة العربية) مقارنةً بالصحافة المصرية والمغربية. وحتى نحيد عن تأثير الحجم على الرغم من محدوديته أو عدم تأثيره كما رأينا آنفا، نختار عشوائيا من موضوعات الصحافة المصرية نفس الحجم نسبةً مع موضوعات الصحافة المغربية. وقد كان عدد الكلمات المميزة ٩٢٢ كلمة عند المقارنة بالصحافة المصرية



و١٠٤٤ كلمة عند المقارنة بالصحافة المغربية.

الكلمة	رتبة الكلمة المميزة بحسب حجم المدونة المرجعية					
	٦,٦٢	٠,٢٥	٠,٥	١	٢	٤
الفريق	١	١	١	١	١	١
مباراة	٢	٣	٣	٣	٢	٢
فريق	٣	٢	٢	٢	٣	٣
المباراة	٤	٤	٤	٤	٤	٤
كرة	٥	٥	٥	٥	٥	٥
المنتخب	٦	٦	٦	٦	٦	٦
القدم	٧	٩	٧	٨	٧	٧
النادي	٨	٧	٨	٧	٨	٨
اللاعب	٩	١٠	١٠	١٠	١٠	١٠
نادي	١٠	٨	٩	٩	٩	٩
الدوري	١١	١٢	١١	١١	١١	١١
كاس	١٢	١٣	١٢	١٢	١٢	١٢
المريخ	١٣	١٨	١٧	١٥	١٥	١٥
اللاعبين	١٤	١٦	١٤	١٤	١٣	١٣
لاعب	١٥	١٧	١٦	١٣	١٤	١٤



الكلمة	رتبة الكلمة المميزة بحسب حجم المدونة المرجعية					
ملعب	١٦	٢٠	١٨	١٩	١٩	١٦
الموسم	١٧	١٤	١٣	١٨	١٨	١٧
الفوز	١٨	١٩	١٥	١٦	١٧	١٩
الاهلي	١٩	١٥	١٩	١٧	١٦	١٨
دوري	٢٠	٢١	٢٠	٢٢	٢٠	٢٠

الجدول (١٣): مقارنة رتبة الكلمات الأكثر تميزا عندما كان حجم المدونة المرجعية أكبر بمقدار ٦,٦٢ من حجم مدونة الدراسة مع بقية الأحجام الأخرى المستخدمة للمدونة المرجعية.

وعلى الرغم من تساوي حجم المدونة المرجعية للصحافة المصرية والمدونة المرجعية للصحافة المغربية إلا أن عدد الكلمات المميزة كان أكثر بالنسبة للصحافة المغربية بمقدار ١٢٢ كلمة، وقد يعود السبب في هذا إلى تقارب المفردات والأساليب بين الصحافة السعودية والصحافة المصرية.

وعند مقارنة الكلمات المميزة في الحالتين، كان عدد الكلمات المميزة المشتركة ٥٩٧ كلمة، أي حوالي ٦٥٪ من الكلمات المميزة عندما كانت المدونة المرجعية هي الصحافة المصرية. ولكن عند النظر للمائتي كلمة الأكثر تميزا نجد أن عدد الكلمات المميزة المشتركة هي ١١٩ كلمة أي ما نسبته ٦٠٪ تقريبا، وهذه النسبة أقل بكثير من حالات التشارك التي رأيناها في تجارب حجم المدونة، وهذا المؤشر يعود تأثيره بذلك إلى تأثير المنطقة الجغرافية على الكلمات المميزة.

ولو نظرنا للكلمات الأربعين الأكثر تميزا عندما كانت المدونة المرجعية هي الصحافة المغربية وقارنا رتبتها بالكلمات المميزة عندما كانت المدونة المرجعية هي الصحافة المصرية كما هو مبين في الجدول (١٤) فسنجد أن هناك تقاربا في الرتبة بين الكلمات المشتركة، غير أن الملحوظ هو غياب بعض الكلمات عندما كانت المدونة المرجعية الصحافة المصرية والكلمات المميزة: (الدكتور، والسعودي، ومحافظة، والطلاب، والمهندس، والنصر،



والهلال، والطالبات، وسلطان، والحوثية). وهذه الكلمات شائعة في الصحافة المصرية والسعودية فلم تظهر ككلمات مميزة عندما كانت المدونة المرجعية الصحافة المغربية. لكن الملاحظ في الكلمات المميزة هي أنها في أغلبها أسماء أعلام مثل (عبدالعزيز، عبدالله، سلمان، الرياض، مكة، الأحساء)، أو ألقاب سياسية واجتماعية وعلمية مثل (الدكتور، الأمير، سمو، خادم، المهندس). ومثل هذه الكلمات -من النظرة الأولى على الأقل- قد لا تعكس اختلافا لغويا بقدر ما تعكس الاختلافات الاجتماعية والسياسية.

ومن أجل أن تكون الصورة أكثر وضوحا: انظر الجدول (١٥) الذي يوضح الكلمات الأربعين الأكثر تميزا عندما كانت الصحافةُ المرجعية الصحافةَ المصرية. ونجد أن الكلمات (المملكة، الملك، الملكي) مميزة للصحافة السعودية عند مقارنتها بالصحافة المصرية لاختلاف النظام السياسي (ملكي مقابل جمهوري) ولا توجد ضمن الكلمات المميزة للصحافة السعودية عند مقارنتها بالصحافة المغربية لتشابه نظامي الحكم. ولكننا نجد بعض الاختيارات في رسم الكلمات أو التفضيلات اللغوية مثل (الشؤون، الملتقى، الجائزة). وهذا بحاجة لدراسة أعمق ليس هنا مكانها. والخلاصة أن للمنطقة الجغرافية تأثيرًا على الكلمات المميزة، فيمكن أن تظهر الاختلافات السياسية والاجتماعية وتبين بعض الفروقات في التفضيلات اللغوية.

الكلمة	رتبة الكلمة		الكلمة	رتبة الكلمة	
	المغرب	مصر		المغرب	مصر
بن	1	1	المكرمة	21	21
عبدالعزيز	2	3	عبدالرحمن	22	27
الدكتور	3	*	فيصل	23	23
الحرمين	4	6	الطلاب	24	*
الامير	5	5	جازان	25	25
الشريفين	6	7	المنورة	26	26

الكلمة	رتبة الكلمة		الكلمة	رتبة الكلمة	
	المغرب	مصر		المغرب	مصر
عبدالله	7	12	عسير	27	24
الرياض	8	10	نجران	28	29
خادم	9	9	نايف	29	28
سمو	10	8	المهندس	30	*
سلمان	11	11	النصر	31	*
فهد	12	13	الهلال	32	*
السعودي	13	*	مبي	33	30
سعود	14	15	الدمام	34	32
السمو	15	14	منسوبي	35	31
محافظة	16	*	الطالبات	36	*
جدة	17	16	سلطان	37	*
ولي	18	18	الاحساء	38	33
معالي	19	19	الطائف	39	38
مكة	20	20	الحوثية	40	*

الجدول (١٤): الكلمات الأربعين الأكثر تميزا للصحافة السعودية عند مقارنتها بالصحافة المغربية ومقارنتها برتبتها عندما كانت المدونة المرجعية هي الصحافة المصرية.

الكلمة	رتبة الكلمة		الكلمة	رتبة الكلمة	
	المغرب	مصر		المغرب	مصر
بن	1	1	المكرمة	21	21
المملكة	2	*	الشؤون	22	*
عبدالعزيز	3	2	فيصل	23	23
الملك	4	*	عسير	24	27
الامير	5	5	جازان	25	25
الحرمين	6	4	المنورة	26	26
الشريفين	7	6	عبدالرحمن	27	22
سمو	8	10	نايف	28	29
خادم	9	9	نجران	29	28
الرياض	10	8	مبي	30	33
سلمان	11	11	منسوبي	31	35
عبدالله	12	7	الدمام	32	34
فهد	13	12	الاحساء	33	38
السمو	14	15	١٤٣٦	34	42
سعود	15	14	الملتقى	35	*
جدة	16	17	تبوك	36	43



الكلمة	رتبة الكلمة		الكلمة	رتبة الكلمة	
	المغرب	مصر		المغرب	مصر
الملكي	17	*	الجائزة	37	*
ولي	18	18	الطائف	38	39
معالي	19	19	الباحة	39	41
مكة	20	20	الله	40	45

الجدول (١٥): الكلمات الأربعين الأكثر تميزا للصحافة السعودية عند مقارنتها بالصحافة المصرية ومقارنتها برتبتها في المدونة المرجعية: الصحافة المغربية.

٣. ٣. ٣. تأثير الموضوع

نكشف في هذا المبحث عن تأثير موضوع المدونة المرجعية على الكلمات المميزة. وللقيام بهذا سنقارن الجزء الخاص بالمواضيع السياسية في الصحافة العربية ٢٠١٥م مع جزء قريب منه وهو المواضيع الدولية، وجزء آخر بعيد عنه وهو الجزء الخاص بالمواضيع الدينية. وسنحيد عامل الحجم ونساوي بين أحجام المدونات لتكون مدونتي المواضيع السياسية والدولية بنفس حجم المواضيع الدينية (بقية العوامل تم تحييدها فالمناطق الجغرافية واحدة والزمن واحد). وتوضح نتائج التجارب أن عدد الكلمات المميزة كان ٣٤٤ كلمة و١٠٧٩ كلمة في المدونة المرجعية: المواضيع الدولية والمواضيع الدينية على التوالي، أي أن عدد الكلمات المميزة في حالة استخدام المواضيع الدينية كمدونة مرجعية أكبر بحوالي ثلاث مرات. وهذا يشير إلى أنه إذا كانت المدونة المرجعية قريبة في موضوعها من مدونة الدراسة فإن عدد الكلمات المميزة يقل والعكس صحيح. وكان عدد الكلمات المميزة المشتركة ٢٠٠ كلمة أي ما نسبته ٥٨٪ من الكلمات المميزة عندما كانت المدونة المرجعية هي المواضيع الدولية. وكان عدد الكلمات المشتركة بين الكلمات المائتين الأكثر تميزا هو ٢٥ كلمة أي ما نسبته ١٢,٥٪ وهي بلا شك نسبة



متدنية جدا عما رأيناه في تجارب تأثير الحجم وتجارب تأثير المنطقة الجغرافية.

الكلمة	رتبة الكلمة		الكلمة	رتبة الكلمة	
	الدولية	الدينية		الدولية	الدينية
التربية	1	*	الدستوري	21	84
D	2	33	الاساتذة	22	*
الازهر	3	*	التلاميذ	23	149
النقابة	4	34	تواضروس	24	103
القضاة	5	*	-[	25	*
العمومية	6	40	الفنانات	26	145
القيامة	7	*	التربوية	27	*
:.	8	67	الذكورية	28	192
القوائم	9	63	الفنانين	29	122
الشغل	10	64	الانبا	30	151
محلب	11	*	المعلمين	31	*
دارفور	12	60	الجهوية	32	176
]	13	*	المنيا	33	215
الري	14	97	المرقسية	34	178
نقابة	15	73	السد	35	111
الاقباط	16	96	الجلالة	36	*

الكلمة	رتبة الكلمة		الكلمة	رتبة الكلمة	
	الدولية	الدينية		الدولية	الدينية
المجيد	17	102	الامتحانات	37	144
ص	18	*	المجتمعى	38	341
المائية	19	81	المجتمعي	39	*
بحيرى	20	*	الهنيدى	40	234

الجدول (١٦): الكلمات الأربعين الأكثر تميزا للمواضيع السياسية عند مقارنتها بالمواضيع الدولية ومقارنتها برتبتها عندما كانت المدونة المرجعية هي المواضيع الدينية.

ولو نظرنا للكلمات الأربعين الأكثر تميزا عندما كانت المدونة المرجعية هي المواضيع الدولية كما هو موضح في الجدول (١٦) فإننا نجد أن ٢٦ كلمة منها من ضمن الكلمات المميزة عندما كانت المواضيع الدينية هي المدونة المرجعية، لكن هناك اختلافا كبيرا في رتبتها، كما أننا لا نجد كلمات تدل دلالة واضحة على الموضوع السياسي، بل تدل بعض الكلمات على الشأن المصري والقبطي منه بشأن خاص. ولو نظرنا للكلمات الأربعين الأكثر تميزا عندما كانت المواضيع الدينية هي المدونة المرجعية كما هو موضح في الجدول (١٧)، فإننا نلاحظ أن المشترك منها هو ثلاث كلمات فقط وبقية الكلمات لم تظهر أبدا كما نجد أن لأغلب الكلمات علاقةً بالمواضيع السياسية والأحداث التي جرت خلال جمع هذه النصوص المتعلقة بـ (عاصفة الحزم).

الكلمة	رتبة الكلمة		الكلمة	رتبة الكلمة	
	الدينية	الدولية		الدولية	الدينية
الانتخابات	1	*	الشعبية	٢١	❋
الحكومة	2	*	تونس	٢٢	❋

الكلمة	رتبة الكلمة		الكلمة	رتبة الكلمة	
	الدينية	الدولية		الدولية	الدينية
الحزب	3	*	اللواء	23	*
الوطني	4	*	ولاية	24	*
حزب	5	*	الجنوب	25	*
السودان	6	*	البشير	26	*
٪	7	*	الامنية	27	*
القوات	8	*	التحالف	28	*
قوات	9	*	الشرطة	29	*
الجزائر	10	*	المغربية	30	*
الحوثيين	11	*	السودانية	31	*
عدن	12	*	القطاع	32	*
المتحدة	13	*	D	33	2
الاحزاب	14	*	النقابة	34	4
الخرطوم	15	*	تعديل	35	*
الحوثي	16	*	الاقتراع	36	*
الانتخابية	17	*	البرلمان	37	*
هادي	18	*	السوداني	38	*

الكلمة	رتبة الكلمة		الكلمة	رتبة الكلمة	
	الدينية	الدولية		الدولية	الدينية
الاتحاد	19	*	مارس	39	*
المعارضة	20	*	العمومية	40	6

الجدول (١٧): الكلمات الأربعين الأكثر تميزا للمواضيع السياسية عند مقارنتها بالمواضيع الدينية ومقارنتها برتبتها عندما كانت المدونة المرجعية هي المواضيع الدولية.

الانتخابات	حزب	الحزب	الأحزاب	الحوثيين	الحوثي	المعارضة	البشير
الانتخابية	التحالف	هادي	الدستور	الاقتراع	الدستورية	البرلمان	دارفور
احزاب	تحالف	انتخابات	الديمقراطي	البرلمانية	الدستوري	الحريات	النواب
لحوثيين	الجبهة	الحوثيون	الناخبين	السودانيين	البابا	الاقباط	القضاة
:.	تواضروس	كردفان	القوائم	المخلوع	المفوضية	غارات	الضالع
الحوثية	الاشتراكي	القيادي	نكيران	مسلحي	السبسي	صعدة	دستورية
المصالحة	الانتقالية	الانبا	المرقسية	الحزبية	بوتفليقة	مليشيات	البوليساريو
النقابات	الثورية	الانتخابي	دستور	التشريعية	عمامرة	عسكري	النقابية
D	مساهل	شبوة	بطريرك	الكاتدرائية	كيران	مزوار	-*
الغارات	حاح	قداس	الرئاسية	المتمردين	غندور	النقابي	الميليشيات

الجدول (١٨): الكلمات الثمانين الأكثر تميزا للمواضيع السياسية في مدونة الصحف العربية عند مقارنتها بجميع مواضيع المدونة ماعدا المواضيع الدولية.

ولكن هل من الأفضل أن تكون المدونة المرجعية ذات موضوع واحد أم ذات مواضيع متعددة؟ ولمعرفة ذلك استخدمت المواضيع السياسية كمدونة دراسة وبقية



مواضيع مدونة الصحافة العربية ٢٠١٥م ما عدا المواضيع الدولية كمدونة مرجعية. وقد كان عدد الكلمات المميزة ١٠٧٨ كلمة مميزة. ويوضح الجدول (١٨) الكلمات الثمانين الأكثر تميزا منها. ويتضح من الجدول عند مقارنته بالجدولين (١٦) و(١٧) أن الكلمات المميزة، وبعد استخدام أكثر من موضوع واستبعاد المواضيع القريبة جدا من مدونة الدراسة، كانت أكثر تعبيرا عن الموضوع، وراصدة للأحداث السياسية خلال فترة صدور النصوص (عاصفة الحزم) ولن أتطرق إلى تحليل هذه الكلمات فهذا ليس مجاله وأترك للقارئ الكريم النظر والتمعن فيها. ونلاحظ أيضا ورود رموز وأحرف إنجليزية (:/ . / ، / -/ */ D) من ضمن الكلمات المميزة، وسأتطرق لطريقة التعامل مع مثل هذه الحالات في نهاية الفصل، فحذفها أو تركها يعتمد على الهدف من استخراج الكلمات المميزة. ولوحظ أيضا كلمة «حاح» الموجودة في السطر الأخير من الجدول وهي نتيجة للتجزيء الخاطئ من قبل برنامج المجزئ العربي. والأصل هو «بحاح» وهو اسم العائلة لرئيس الوزراء اليمني في وقتها.

وما يمكن الخلوص إليه، بشكل عام، من التجارب الخاصة بموضوع المدونة المرجعية هو أن لموضوعها أثرًا كبيرًا في عدد الكلمات المميزة ونوعيتها، ومن المفضل أن تكون متعددة الموضوعات وبعيدة عن موضوع مدونة الدراسة.

٣. ٣. ٤. تأثير الزمن

لدراسة تأثير اختلاف الفترات الزمنية على نتائج الكلمات المميزة بين المدونة الرئيسة (مدونة الدراسة) والمدونة المرجعية، نستخرج-على سبيل المثال-الكلمات المميزة لموضوع الرياضة من مدونة الصحافة السعودية لعام ٢٠٠٧م وذلك عن طريق مقارنتها بالمواضيع الأخرى في المدونة نفسها، ثم نقارنها في مرحلة ثانية بمدونة الصحف السعودية للعام ٢٠١٢م، وفي مرحلة ثالثة بالصحافة السعودية من مدونة الصحف العربية (٢٠١٥م). وبهذا يتم تحييد الاختلافات الجغرافية تماما وكذلك الاختلاف الذي قد يحصل نتيجة لوعاء النص. وبالنسبة لاختلاف موضوعات المدونة المرجعية نُبقي فقط على الموضوعات المشتركة بين المدونات الثلاث وهي: الاجتماعية، والثقافية، والاقتصادية، والسياسية وبذلك يتم تحييد الاختلافات حول موضوعات المدونة المرجعية. ولتحييد تأثير حجم المدونة المرجعية فقد كانت جميع المدونات وموضوعاتها متساوية، وكانت



أحجامها متساوية لما هو موجود في مدونة الصحافة السعودية للعام ٢٠١٢م؛ لأنها الأصغر، وكانت أحجامها أكبر بخمس مرات من حجم مدونة الدراسة. وكما حصل في التجارب السابقة فقد تم اختبار النصوص بصورة عشوائية أيضا، وبهذا نكون قد حيّدنا جميع المؤثرات الممكنة بقدر الإمكان وأبقينا على تأثير الفترات الزمنية فقط.

وقد كان عدد الكلمات المميزة ٥٢٤ كلمة، و٥١٨ كلمة، و٥٧٧ كلمة للحالات الثلاث المذكورة أعلاه على التوالي. والفرق بين عدد الكلمات المميزة معدوم تقريبا (٦ كلمات) بين عامي ٢٠٠٧م و٢٠١٢م، وقليل بين عام ٢٠٠٧م و٢٠١٥م (٥٣ كلمة). وقد كان عدد الكلمات المميزة المشتركة، عندما كانت المدونة المرجعية هي موضوعات مدونة الصحافة السعودية عام ٢٠٠٧م وعندما كانت المدونة المرجعية هي مدونة الصحافة السعودية، ٤٥٤ كلمة (أي ما نسبته ٨٧٪)، وعدد الكلمات المشتركة بينهما من الكلمات المئتين الأكثر تميزا ١٧١ كلمة (أي ما نسبته ٥,٨٦٪)، وكان عدد الكلمات المميزة المشتركة، عندما كانت المدونة المرجعية تتضمن موضوعات مدونة الصحافة السعودية للعام ٢٠٠٧م وعندما كانت المدونة المرجعية هي الصحافة السعودية للعام ٢٠١٥م، ٤٤٢ كلمة (أي ما نسبته: ٨٤٪)، وعدد الكلمات المشتركة بينهما من الكلمات المئتين الأكثر تميزا ١٦٤ كلمة (أي ما نسبته: ٨٢٪)، ولكن: الملاحظ أن الكلمات المشتركة الأكثر تميزا تحتفظ بنفس رتبتها تقريبا مع اختلاف الفترات الزمنية كما يوضح ذلك الجدول (١٩). وتشير هذه البيانات إلى أن تأثير اختلاف الفترات الزمنية موجود ولكنه قليل مثلما هو الحال في تأثير المناطق الجغرافية.

الكلمة	زمن صدور النصوص			الكلمة	زمن صدور النصوص		
	2007	2012	2015		2007	2012	2015
الفريق	1	1	5	اللاعبين	16	14	12
كرة	2	4	3	النصر	17	13	18
المنتخب	3	5	2	مدرب	18	17	17



الكلمة	زمن صدور النصوص			الكلمة	زمن صدور النصوص		
	2007	2012	2015		2007	2012	2015
المباراة	4	3	4	منتخب	19	21	19
الاتحاد	5	2	1	الشوط	20	20	20
فريق	6	7	7	البطولة	21	18	24
الهلال	7	6	10	الكرة	22	19	21
القدم	8	8	6	لاعب	23	22	22
نادي	9	16	13	لاعبي	24	24	23
كاس	10	9	8	بطولة	25	23	27
مباراة	11	10	9	الدوري	26	27	28
المدرب	12	11	16	مباريات	27	25	26
النادي	13	*	15	دوري	28	26	31
اللاعب	14	12	11	الموسم	29	29	30
الاهلي	15	15	14	فهد	30	*	*

الجدول (١٩): الثلاثون كلمة الأكثر تميزا لموضوع الرياضة عند استخدام مدونة الصحافة السعودية لعام ٢٠٠٧ ومقارنة رتبتها عندما كانت المدونة المرجعية هي الصحافة السعودية عام ٢٠١٢م وعام ٢٠١٥م.

٣. ٤. تأثر المقاييس الإحصائية

أوضحتُ في المبحث السابق تأثير المدونة المرجعية على الكلمات المميزة مستخدما في ذلك واحدا من المقاييس الإحصائية الأكثر استخداما: معامل الاحتمال اللوغاريثمي. ولكن هل ما ينطبق على هذا المقياس ينطبق على المقاييس الأخرى؟ أولُ ما يجب الانتباه



إليه هنا هو أن نتائج تطبيق كل مقياس تكون مختلفة في حال تطبيقها من المقاييس الأخرى، فالكلمات المميزة وعددها ورتبها سوف تتباين بلا شك بين المقاييس، ولكننا هنا وبسبب ضيق المساحة سوف نقتصر على دراسة الاختلاف والتشابه بين المقاييس الإحصائية عند اختلاف حجم المدونة المرجعية، والكلمات الأكثر تميزا عند استخدام كل معيار.

وقد أَجريتُ نفس التجارب مثل ما هو موضح في المبحث (٣. ١.) باستخدام جميع المقاييس الإحصائية. وهنا يجب التنويه إلى أمرين مهمين؛ الأول: هو أن تطبيق جميع المحددات مثلما هو الحال بالنسبة لمعامل الاحتمال اللوغاريثمي لم يكن ممكنا بالنسبة لمعامل تي، فقد كان أقل نسبة خطأ أمكن الحصول عليها للكلمات المميزة، عندما كان حجم المدونة المرجعية هو ربع مدونة الدراسة، ما نسبته ٩٨٪، أما الثاني: فهو أن كلاً من معامل الغرابة والمعلومات المتبادلة ليست لهما نسبة ثقة، وبالتالي تم احتساب جميع الكلمات التي تحقق شرط نسبة الكلمة في مدونة الدراسة إلى المدونة المرجعية وشرط تكرار الكلمة في مدونة الدراسة بشكل مساوٍ لـ أو أكبر من ١٠، ثم تم ترتيب الكلمات بحسب قيمة المقياس الإحصائي. ويوضح الجدول (٢٠) عدد الكلمات المميزة المستخرجة من المواضيع الرياضية في مدونة الصحف العربية لكل مقياس من المقاييس الإحصائية الموضحة في الجدول (٧).

وما نلحظه من البيانات الموضحة في الجدول (٢٠) هو أن مربع كاي ومعامل زد ومعامل تي لهما نفس سلوك الاحتمال اللوغاريثمي حيث يزداد عدد الكلمات المميزة مع ازدياد حجم المدونة المرجعية ولكن معامل الغرابة والمعلومات المتبادلة لهما سلوك معاكس حيث يقل عدد الكلمات المميزة مع ازدياد حجم المدونة المرجعية. وألفت نظر القارئ الكريم إلى ملاحظتين مهمتين؛ الأولى: هي أن عدد الكلمات المميزة عند استخدام معامل الغرابة هو نفس عدد الكلمات المميزة عند استخدام المعلومات المتبادلة، وهذا طبيعي بسبب عدم وجود قيمة حرجة لنتيجة تطبيق المعيار الإحصائي، نقبل عن طريقها ما يساويها أو أكبر منها مثلما هو الحال لبقية المقاييس الإحصائية. ولكن، عند اختيار الكلمات الأكثر تميزا من المحتمل جدا اختلاف الكلمات المميزة حيث يتم اختيار الكلمات ذات القيمة الأعلى بحسب المعيار الإحصائي؛ والملاحظة الثانية أن عدد الكلمات المميزة متساوٍ تقريبا مع جميع المقاييس الإحصائية ماعدا معامل تي عندما كان حجم المدونة المرجعية يساوي ٦,٦٢ مرة حجمَ مدونة الدراسة. وهذا



يؤيد أن حجم المدونة المرجعية على عدد الكلمات يتضاءل عندما يكون حجم المدونة المرجعية أكبر من مدونة الدراسة بخمس مرات أو أكبر.

النسبة بين المدونتين	الاحتمال اللوغاريثمي	مربع كاي	معامل زد	معامل تي	معامل الغرابة	معلومات متبادلة
٠,٢٥	١,٠٩٠	٧١٧	١٨١	١٣	٤,١٧٤	٤,١٧٤
٠,٥٠	١,٦٣٧	١,١٧٠	٥١٧	٥٢	٣,٧١٠	٣,٧١٠
١,٠٠	٢,٣٠١	١,٨٤٧	١,١٩٣	١٢١	٣,٥٤٧	٣,٥٤٧
٢,٠٠	٢,٩١٣	٢,٦٩٠	٢,١٢٨	٢١٦	٣,٤٧٧	٣,٤٧٦
٤,٠٠	٣,٢٤٢	٣,٣٥٧	٣,٢١٣	٢٧٦	٣,٤٢٥	٣,٤٢٤
٦,٦٢	٣,٣١٩	٣,٣٥٨	٣,٣٤٤	٣٥٢	٣,٣٤٤	٣,٣٤٤

الجدول (٢٠): عدد الكلمات المميزة مقابل حجم المدونة المرجعية بالنسبة لمدونة الدراسة.

وقد اشتركت جميع المقاييس الإحصائية ماعدا معامل الغرابة في ظاهرة ازدياد قيمة المعامل الإحصائي للكلمات المميزة مع ازدياد حجم المدونة المرجعية كما هو مبين في الجدول (٢١) الذي يوضح قيم المعاملات الإحصائية للكلمة الأولى بالترتيب عند استخدام معامل الاحتمال اللوغاريثمي في جميع تجارب تأثير حجم المدونة. على سبيل المثال: كلمة «الفريق». أما قيم معامل الغرابة فليس لها نمط ثابت بالنسبة للازدياد أو الانخفاض، فهي تزداد مرة وتنخفض أخرى بمقدار بسيط عن سابقتها في الحجم، ونستطيع القول إنها ثابتة تقريبا. ويُشير ازدياد قيم المعاملات بدرجة كبيرة إلى قلة نسبة الخطأ وازدياد نسبة الثقة في النتائج المستخرجة مع ازدياد حجم المدونة.

القيمة الحرجة	0.25	0.50	1.00	2.00	4.00	6.62
الاحتمال اللوغاريثمي	4,531.3	8,354.1	13,550.5	20,220.0	26,793.3	35,776.1



مربع كاي	2,752.6	5,517.1	10,569.9	19,024.9	31,688.9	61,135.7
معامل زد	23.4	42.8	72.3	110.3	154.6	230.4
معامل تي	21.0	35.1	52.0	67.5	79.5	91.1
معامل الغرابة	37.9	48.7	37.9	42.8	39.3	34.6
معلومات متبادلة	0.3	0.6	0.9	1.4	1.9	2.7

الجدول (٢١): قيم المعاملات الإحصائية لكلمة «الفريق» مع ازدياد حجم المدونة المرجعية.

المعيار الإحصائي	حجم المدونة المرجعية بالنسبة لمدونة الدراسة				
	0.5	1	2	4	6.62
الاحتمال اللوغاريثمي	89.7%	91.0%	91.6%	91.2%	90.8%
مربع كاي	93.0%	94.0%	94.0%	93.0%	90.0%
معامل زد	94.5%	94.5%	95.0%	92.8%	91.2%
معامل تي	92.3%	92.3%	92.3%	92.3%	92.3%
معامل الغرابة	58.5%	44.5%	32.5%	21.5%	10.5%
معلومات متبادلة	58.5%	44.5%	32.5%	21.5%	10.5%

الجدول (٢٢): نسبة التشارك في المائتي كلمة الأكثر تميزا بحسب المعيار الإحصائي وحجم المدونة المرجعية اعتمادا على النتائج عندما كان حجم المدونة المرجعية مساويا ٢٥ ,٠ من حجم مدونة الدراسة.



ولكن ما نسبة التشارك في الكلمات الأكثر تميزا والمستخرجة باستخدام الطرق الإحصائية المختلفة؟ يُوضح الجدول نسبةَ التشارك في المائتي كلمة الأكثر تميزا عندما كان حجم المدونة المرجعية يساوي ٠,٢٥ مقارنة بحجم مدونة الدراسة مع بقية الكلمات المميزة المستخرجة من الأحجام المختلفة للمدونة المرجعية لجميع المقاييس الإحصائية المستخدمة. وفي حالة معامل تي ومعامل زد كان عدد الكلمات المميزة ١٣ و١٨١ كلمة على التوالي، فاعتبرناها المرجع وقسنا عليها بدلا من المئتين. وكما توضح البيانات فإن نسبة التشارك (الجدول ٢٢) عالية بالنسبة لمربع كاي ومعامل زد ومعامل تي مثلما كان عليه الحال بالنسبة لجميع الأحجام المستخدمة للمدونة المرجعية. أما بالنسبة لمعامل الغرابة والمعلومات المتبادلة فإن نسبة التشارك ضعيفة جدا، وتقل مع ازدياد حجم المدونة المرجعية، كما أن نسبة التشارك متساوية تماما بالنسبة للمعيارين (يعود هذا لأنهما يعطيان نفس الكلمات المتميزة كما سوف نرى لاحقا).

ولدراسة مدى التشابه بين المقاييس الإحصائية واختلافها في استخراجها للكلمات المميزة، سوف أقيس نسبة التشارك في المائتي كلمة الأكثر تميزا بين جميع المقاييس عندما كان حجم المدونة المرجعية يساوي ٦,٠٢ من حجم الدراسة. وكما هو موضح في الجدول (٢٣) فإن الكلمات الأكثر تميزا تكاد تتطابق (٩٩,٥٪) بين معامل الاحتمال اللوغاريثمي ومربع كاي ومعامل زد، ونتائج هذه المقاييس الثلاث تتشابه إلى حد كبير جدا مع معامل تي (٩٤,٥٪)، مما يعني تشابها كبيرا في النتائج بين هذه المقاييس. كما تتطابق نتائج معامل الغرابة مع المعلومات المتبادلة ولكنها تختلف تماما عن نتائج كل المقاييس الإحصائية الأربعة الأخرى حيث تنحصر نسبة التشابه بين ٢,٥ و١.

	الاحتمال اللوغاريثمي	مربع كاي	معامل زد	معامل تي	معامل الغرابة	معلومات متبادلة
الاحتمال اللوغاريثمي	100.0%	99.5%	99.5%	94.5%	2.5%	2.5%
مربع كاي		100.0%	100.0%	94.5%	2.0%	2.0%
معامل زد			100.0%	94.5%	2.0%	2.0%

معامل تي				100.0%	1.0%	1.0%
معامل الغرابة					100.0%	100.0%
معلومات متبادلة						100.0%

الجدول (٢٣): نسبة التشارك بين المقاييس الإحصائية في المائتي كلمة الأكثر تميزا عندما كان حجم المدونة المرجعية ٦,٦٢ حجم مدونة الدراسة.

ولكن هل هذا التشابه على أهميته ودلالته يدل أيضا على أن الكلمات المستخرجة لها نفس الرتبة؟ توضح المقارنة بين مراتب الكلمات المميزة ما هو آتٍ:

أولاً: التشابه الكبير في الرّتب عند استخدام معامل الاحتمال اللوغاريثمي ومربع كاي ومعامل تي ومعامل زد، حيث تحتل الكلمات الأكثر تميزا نفس الرّتب تقريبا (انظر جدول ٢٤).

ثانيًا: تحتل الكلمات الأكثر تميزا، عند استخدام المقاييس الأربعة المذكورة آنفا، مراتب متدنية عند استخدام معامل الغرابة والمعلومات المتبادلة (انظر الجدول ٢٤).

ثالثًا: التطابق التام في مراتب الكلمات المميزة عند استخدام معامل الغرابة والمعلومات المتبادلة (انظر الجدول ٢٥).

رابعًا: تحتل الكلمات الأكثر تميزا، عند استخدام معامل الغرابة والمعلومات المتبادلة، مراتب متدنية عند استخدام معامل الاحتمال اللوغاريثمي ومربع كاي ومعامل تي ومعامل زد (انظر الجدول ٢٥).

خامسًا: إن الكلمات الأكثر تميزا والمستخرجة عن طريق معامل الاحتمال اللوغاريثمي ومربع كاي ومعامل تي ومعامل زد هي كلمات تدل مباشرة على موضوع مدونة الدراسة (لاحظ الكلمات الموضحة في الجدول ٢٤ التي تدل مباشرة على الموضوع الرياضي).

سادسًا: إن الكلمات الأكثر تميزا والمستخرجة عن طريق معامل الغرابة والمعلومات المتبادلة عبارة عن أسماء أعلام أو كلمات أجنبية (لاحظ الكلمات الموضحة في الجدول ٢٥ كمثال على ذلك).



الكلمة	الاحتمال اللوغاريثمي	مربع كاي	معامل تي	معامل زد	معامل الغرابة	معلومات متبادلة
الفريق	١	١	١	١	٢٢٠٤	٢٢٠٤
مباراة	٢	٢	٣	٢	١١٥٤	١١٤٧
فريق	٣	٣	٢	٣	٢٥٢٩	٢٥٢٩
المباراة	٤	٤	٤	٤	١١٧٩	١١٧٩
كرة	٥	٥	٥	٥	١٦١٦	١٦١٦
المنتخب	٦	٦	٧	٦	١٤٨٨	١٤٨٨
القدم	٧	٧	٦	٧	١٩٠٠	١٩٠٠
النادي	٨	٨	٨	٨	١٧٨٩	١٧٨٩
اللاعب	٩	١٠	١١	١٠	١١٦١	١١٦١
نادي	١٠	٩	٩	٩	٢٠٠٣	٢٠٠٣
الدوري	١١	١١	١٠	١١	١٧٢٥	١٧٢٥
كاس	١٢	١٢	١٣	١٢	١٤٦٩	١٤٦٩
المريخ	١٣	١٣	١٧	١٣	١٠٩٤	١٠٨٩
اللاعبين	١٤	١٤	١٦	١٤	١١٤٦	١١٤٦
لاعب	١٥	١٥	١٥	١٥	١٣٦٣	١٣٦٣
ملعب	١٦	١٧	١٩	١٧	١٥٥٤	١٥٥٤
الموسم	١٧	١٦	١٢	١٦	٢٢٧١	٢٢٧١



الكلمة	الاحتمال اللوغاريثمي	مربع كاي	معامل تي	معامل زد	معامل الغرابة	معلومات متبادلة
الفوز	١٨	١٨	١٤	١٨	١٩٠١	١٩٠١
الاهلي	١٩	١٩	١٨	١٩	١٨٠٨	١٨٠٨
دوري	٢٠	٢٠	٢٢	٢٠	١٥١١	١٥١١

الجدول (٢٤): العشرون كلمة الأكثر تميزا عند استخدام معامل الاحتمال اللوغاريثمي وعندما كان حجم المدونة المرجعية هو ٦٢, ٦ من حجم مدونة الدراسة ومقارنتها برتبة نفس الكلمات عند استخدام المقاييس الأخرى.

الكلمة	معامل الغرابة	معلومات متبادلة	الاحتمال اللوغاريثمي	مربع كاي	معامل تي	معامل زد
غارزيتو	١	١	١١٤	١٢٢	١٤٣	١٢٢
الصفاقسي	٢	٢	١٣٤	١٤٠	١٥٨	١٤٠
الكوكي	٣	٣	١٧٥	١٨١	٢٠٩	١٨١
كورابيا	٤	٤	١٨٩	١٩١	٢٢٦	١٩١
البنزرتي	٥	٥	١٩٩	٢٠٤	٢٣٨	٢٠٤
modified	٦	٦	٢٣٩	٢٤٧	٢٩٠	٢٤٧
غوركوف	٧	٧	٢٥٢	٢٦٥	٣٠٠	٢٦٥
تشافي	٨	٨	٢٦٨	٢٧١	٣١٢	٢٧١
الليغا	٩	٩	٣٢٤	٣٣١	*	٣٣١
1-Mar	١٠	١٠	٣٤٩	٣٦٣	*	١١٣٤



الكلمة	معامل الغرابة	معلومات متبادلة	الاحتمال اللوغاريثمي	مربع كاي	معامل تي	معامل زد
البافاري	١١	١١	٣٦٠	٣٦٨	*	٣٦٨
مباراتي	١٢	١٢	٣٧٢	٣٨١	*	٣٨١
السيتي	١٣	١٣	٣٧٣	٣٨٢	*	٣٨٢
ركنية	١٤	١٤	٣٧٧	٣٨٦	*	٣٨٦
بلفوضيل	١٥	١٥	٣٩٩	٤٠٥	*	٤٠٥
سكورز	١٦	١٦	٤٠٣	٤٠٨	*	٤٠٨
فياريال	١٧	١٧	٤٠٥	٤١٥	*	٤١٥
البرسا	١٨	١٨	٤٠٧	٤٢٠	*	٤٢٠
المنستيري	١٩	١٩	٤١٢	٤٢٦	*	٤٢٦
Jan-00	٢٠	٢٠	٤١٩	٢٢٥٤	*	٢٢٥٤
الجرجيسي	٣٩	٣٩	٦٣٣	٦٤٧	*	٦٤٧
فيلايني	٤٠	٤٠	٦٣٤	٦٤٨	*	٦٤٨

الجدول (٢٥): العشرون كلمة الأكثر تميزا عند استخدام معامل الغرابة وعندما كان قيمة المدونة المرجعية الحرجة ٦,٦٢ من حجم مدونة الدراسة ومقارنتها بمرتبة نفس الكلمات عند استخدام المقاييس الأخرى.

٣. ٥. ماذا بعد استخراج الكلمات المميزة؟

بعد استخراج الكلمات المميزة هناك أمران مهمان ينبغي للباحث تقريرهما، وهما على النحو الآتي: هل يدرس الباحث جميع ما نتج من عملية استخراج الكلمات المميزة أم أنه يستطيع استبعاد بعض الأصناف من الكلمات المميزة؟ ما عدد الكلمات المميزة التي



سوف يختارها من قائمة الكلمات المميزة بعدما استبعد بعض الأصناف التي لا تخدم دراسته؟

٣ . ٥ . ١ . ما الذي يمكن أن يستبعده الباحث من الكلمات المميزة؟

ينصح بعض الباحثين من قبيل جونسون وإنسلين Johnson and Ensslin (٢٠٠٦) بتنقيح الكلمات المميزة عن طريق استبعاد بعض الأصناف من الكلمات إن رأى الباحث أنها لا تخدم دراسته مثل:

١ . الكلمات الوظيفية.

٢ . أسماء الأعلام.

٣ . الاختلافات الإملائية مثل كلمة مسئول ومسؤول أو الاختلافات التصريفية مثل مظاهرة وتظاهرة. وقبل الاستبعاد يجب النظر في ورود الكلمة واختلافاتها الإملائية أو التصريفية بين المدونتين.

٤ . الكلمات التي تميز السجلات اللغوية الموجودة في جميع مواضيعها مثل صرح، قال، أفاد في لغة الصحافة ومثل بحث، ونظرية، ومقدمة في الكتابة الأكاديمية.

٥ . الأخطاء الإملائية والرموز والكلمات الأجنبية.

وبطبيعة الحال فإن ما يمكن استبعاده في دراسة ما قد يكون هو الأهم في دراسة أخرى. فضمائر المتكلم على سبيل المثال قد تكون مهمة جدا في التحليل النقدي للخطاب critical discourse analysis ولكنها قد لا تكون مهمة في التعرف على موضوع مدونة الدراسة، ومثلها أسماء الأعلام كذلك. ويجب أن أُذكّر القارئ الكريم بأن الباحث مطالب بذكر ما استبعده من أصناف وماهي أسباب استبعاده لها.

٣ . ٥ . ٢ . ما عدد الكلمات المميزة التي يمكن أن ندرسها؟

كما رأينا في التجارب السابقة، فإن عملية استخراج الكلمات المميزة ينتج عنها في العادة عدد كبير من الكلمات مما يجعل دراستها جميعا أمرا قد يكون مستحيلا، ولذلك يلجأ الباحثون إلى اختيار الكلمات الأكثر تميزا بحسب المعيار الإحصائي المستخدم وذلك بعد استبعاد بعض أصناف الكلمات التي لا تخدم البحث، مثل اختيار الخمسين أو المئة أو المئتين كلمة الأولى في القائمة. وقد يلجأ البعض إلى اختيار نسبة معينة مثل ٥٪ أو ١٠٪ إلخ. وفي الحقيقة فإنه لا يوجد حد معين لهذه المعايير، واختيار العدد



يخضع لطبيعة الدراسة وحدودها. وقد يلجأ بعض الباحثين إلى ضم الكلمات التي لها نفس الجذع من الكلمات المميزة واحتسابها كلمةً واحدة مثل (الكلمات المميزة: الفريق، فريق، الفريقين، تجمع معا وتحتسب كلمة واحدة).

وبعد أن ينهي الباحث عملية اختيار الكلمات المميزة يستطيع أن يدرسها عن طريق الخطوات الثلاث الآتية:

الأولى: تصنيفها في مجموعات دلالية معينة قد تنبئ عن الأنماط العامة لمدونة الدراسة ومدى تنوعها.

الثانية: دراسة دلالات الكلمات المميزة ومعرفة التنوعات لهذه الدلالات عن طريق الكشاف السياقي.

الثالثة: استخدام مقاييس التلازم والتصاحب اللفظي لمعرفة ما يتلازم مع الكلمات المميزة.

٣. ٦. الخاتمة

تطرقتُ في هذا الفصل إلى مفهوم الكلمات المميزة وكيفية استخراجها بالإضافة إلى بعض المقاييس الإحصائية التي يمكن استخدامها في عملية الاستخراج. ثم حللتُ تجريبيا تأثير المدونة المرجعية على الكلمات المميزة من حيث حجمها، وموضوعها، والمنطقة الجغرافية التي صدرت فيها نصوصها، والزمن الذي كتبت فيه. وما توصلتُ إليه من هذه التجارب ألخصه على النحو الآتي:

أولاً: أكثر العوامل تأثيرا على الكلمات المميزة هو موضوع المدونة المرجعية. وهو أول ما يجب على الباحث أن يوليه اهتماما أوليًا. ومن المفضل أن تكون المدونة المرجعية متعددة الموضوعات وبعيدة عن موضوع مدونة الدراسة ولكنها من نفس السجل اللغوي.

ثانيًا: تأثير حجم المدونة المرجعية ضئيل على الكلمات الأكثر تميزا، وأن الميزة الوحيدة لأن يكون حجم المدونة المرجعية أكبر من حجم مدونة الدراسة هو ازدياد عدد الكلمات المميزة واستقرارها. ولكن هذا لا يعني تعمد اختيار مدونة مرجعية أصغر، بل يجب على الباحث الحرص على استخدام مدونة مرجعية أكبر متى ما توفرت وكانت مناسبة لموضوع بحثه. وأن الدراسات المبنية على مدونات مرجعية صغيرة يمكن الوثوق بنتائجها والاعتماد عليها.



ثالثًا: تأثير زمن المدونة المرجعية ومنطقتها الجغرافية قليل مقارنة بتأثير الموضوع، كون الزمن والمنطقة يكشفان عن الفروقات الاجتماعية والثقافية والسياسية إن وجدت، والأحداث التي رافقت كل فترة زمنية.

وقد حللتُ تجريبيا تأثير المقاييس الإحصائية المستخدمة على عدد الكلمات المميزة، ونسبة التشارك في الكلمات الأكثر تميزا ونوعية الكلمات المميزة وما خلصت إليه من هذه التجارب هو:

رابعاً: يُمكن تقسيم المقاييس الإحصائية إلى مجموعتين، تتشابه عناصر كل مجموعة منها في السلوك، وتنتج ذات الكلمات الأكثر تميزا تقريبا وبنفس الرتبة في الغالب. وتضم المجموعة الأولى مقياس الاحتمال اللوغاريثمي، ومربع كاي، ومعامل تي، ومعامل زد. وتضم المجموعة الثانية معامل الغرابة والمعلومات المتبادلة.

خامسًا: تنتج مقاييس الاحتمال اللوغاريثمي ومربع كاي ومعامل تي ومعامل زد كلمات مميزة دالةً على موضوع مدونة الدراسة بشكل جلي. وهذا يدعو للقول إنها أفضل ما يستخدم في الكشف عن الاتجاهات والآراء والمواضيع ودراسات التحليل النقدي للخطاب والتعرف على موضوعات المدونات اللغوية. ولعل معامل الاحتمال اللوغاريتمي هو الأفضل فيها (بحسب ما أرى) لأنه لا يفترض أي توزيع إحصائي للبيانات مثلما هو الحال في المقاييس الثلاثة الأخرى.

سادسًا: أن معامل الغرابة والمعلومات المتبادلة تنتج في الأغلب أسماء الأعلام والأخطاء الكتابية وجميع هذه الكلمات لا ترد أبدا في المدونة المرجعية. هذا يجعل من هذين المقياسين طريقة مفضلة للكشف عن الأخطاء الإملائية أو التركيبية في كتابات متعلمي اللغة أو في فحص المدونات والكشف عن الأخطاء المنتظمة فيها والحاصلة جراء تغيير الترميز أو نقل البيانات.

ولعلي أنوه هنا إلى أنني لم أستوف جميع ما توصلتُ إليه من هذه التجارب، مثل سلوك المقاييس الإحصائية مع اختلاف المدونة المرجعية من حيث الموضوع والزمن والمنطقة الجغرافية. كما أنني كنت أود دراسة مقاييس إحصائية أخرى مثل معامل دايس ولوغاريثم دايس وغيرهما ومدى صلاحيتها في استخراج الكلمات المميزة، ومدى تقاربها مع المجموعتين المذكورتين آنفا. وهناك موضوع آخر قد يكون ذا أهمية بالنسبة للكلمات المميزة وهو توزيع الكلمات في النصوص (أي عدد النصوص التي وردت فيها



الكلمة) على الكلمات المميزة. كل هذه بحاجة إلى دراسة وتحليل لم يتسع لمقام الوقت وحجم الفصل إنجازه وعرضه بالتجارب الشاملة.

قائمة المراجع

أولاً: المراجع العربية

- الشمري، عقيل والثبيتي، عبدالمحسن. (٢٠١٥). «لسانيات المدونات: نماذج وتطبيقات في الصحافة العربية». في: المدونات اللغوية العربية: بناؤها وطرق الإفادة منها. تحرير صالح فهد العصيمي. مركز الملك عبدالله بن عبدالعزيز الدولي لخدمة اللغة العربية، الرياض (المملكة العربية السعودية)، الصفحات ١٧٩-٢٣٣.
- ماكنري، توني وهاردي، أندرو. (٢٠١٦). لغويات المدونة الحاسوبية: المنهج والنظرية والتطبيق. ترجمة سلطان بن ناصر المجيول. دار جامعة الملك سعود، الرياض.
- المجيول، سلطان. (٢٠١٥). «البحث اللغوي في المدونات العربية الحاسوبية بين الممكن والمحتمل والمأمول». في» المدونات اللغوية العربية: بناؤها وطرق الإفادة منها. تحرير صالح بن فهد العصيمي. مركز الملك عبدالله بن عبدالعزيز الدولي لخدمة اللغة العربية، الرياض (المملكة العربية السعودية)، الصفحات ٢٣٥-٢٧٩.

ثانيا: المراجع الأجنبية

- Alba-Juez, L. (2013). Nuria Lorenzo-Dus (ed.) Spanish at Work. Analysing Institutional Discourse across the Spanish-Speaking World. Spanish in Context, 10(1): 137-143.
- Al-Thubaity, A., & Al-Subaie, A. (2015, October). Effect of word segmentation on Arabic text classification. In 2015 International Conference on Asian Language Processing (IALP), IEEE, pp. 127-131.
- Al-Thubaity, A., Khan, M., Al-Mazrua, M., & Al-Mousa, M. (2013). New Language Resources for Arabic: Corpus Containing More Than Two Million Words and a Corpus Processing Tool. In Asian Language Processing (IALP), 2013 International Conference on IEEE, pp. 67-70.



- Baker, P. (2004). Querying Keywords Questions of Difference, Frequency, and Sense in Keywords Analysis. Journal of English Linguistics, 32(4), 346-359.
- Baker, P. (2010). Representations of Islam in British broadsheet and tabloid newspapers 1999–2005. Journal of Language and Politics, 9(2): 310-338.
- Berber-Sardinha, T. (2000). Comparing corpora with WordSmith Tools: How large must the reference corpus be? In Proceedings of the workshop on Comparing Corpora-Volume 9, Association for Computational Linguistics. pp. 7-13.
- Cheng, W. (2011). Exploring corpus linguistics: Language in action. Routledge.
- Goh, G. Y. (2011). Choosing a reference corpus for keyword calculation. Linguistic Research, 28(1): 239-256.
- Grabowski, Ł. (2013). Register variation across English pharmaceutical texts: A corpus-driven study of keywords, lexical bundles and phrase frames in patient information leaflets and summaries of product characteristics. Procedia-Social and Behavioral Sciences, 95: 391-401.
- Granger, S., and Paquot, M. (2013). Language for specific purposes learner corpora. Language for specific purposes. The encyclopedia of applied linguistics. New York: Blackwell.
- Johnson, S., & Ensslin, A. (2006). Language in the news: Some reflections on keyword analysis using WordSmith Tools and the BNC. Leeds Working Papers in Linguistics and Phonetics, 11: 96-109.
- Kilgarriff, A. (2012). Getting to know your corpus. In: International Conference on Text, Speech and Dialogue (pp. 3-15). Springer Berlin Heidelberg.
- Monroe, W., Green, S., & Manning, C. D. (2014). Word Segmentation of Informal Arabic with Domain Adaptation. In ACL (2) (pp. 206-211).
- Scott, M. & Tribble, C., 2006, Textual Patterns: keyword and corpus analysis in language education, Amsterdam: Benjamins.

الفصل الرابع

د. عبدالله بن يحيى الفيفي

المعالجة الآلية لمدونات المتعلمين: تحليل الأخطاء في المدونة اللغوية لمتعلمي اللغة العربية

الملخص

يهدف البحث إلى تسليط الضوء على استخدام المدونات اللغوية للمتعلمين Learner Corpora كمصدر لغوي لمجال تحليل الأخطاء Error Analysis في الدراسات المعاصرة، وآلية الاستفادة منها في هذا النوع من الدراسات؛ فيبدأ البحث بمقدمات أساسية حول هذا المجال، تتضمن المقصود بتحليل الأخطاء، وبيان أهميته، والتفريق بين الغلط والخطأ، مع محاولة لتأصيل نشأة هذا المجال، وبيان مراحل تطوره، والبدايات الحديثة لأبحاث تحليل الأخطاء، كما يشرح البحثُ الأسس النظرية التي قام عليها هذا الميدان، منتقلاً بعد ذلك إلى مناقشة مفصلة للمنهجية العلمية المستخدمة في تحليل الأخطاء، التي تتضمن مجموعة من الخطوات الرئيسة الهادفة بالوصول إلى نتائج عملية يمكن الاستفادة منها وتطبيقها في تعليم اللغة الثانية وتعلمها. ولتحقيق الهدف من البحث فإنه يطبق هذه المنهجية على عينات واقعية من المدونة اللغوية لمتعلمي اللغة العربية، فيقدم بذلك نموذجاً عملياً للاستفادة من مدونات المتعلمين في مجال تحليل الأخطاء.

الكلمات المفتاحية: تحليل الأخطاء، المدونة اللغوية، تعليم اللغة، أخطاء الأداء، أخطاء الكفاية.



٤ . ١ . المقدمة

يبرز موضوع تحليل الأخطاء عند الحديث عن تعليم اللغة الثانية وتعلمها، ليحتل مكاناً واضحاً كأحد المناهج الأساسية في الدراسات الحديثة في هذا المجال، بل إنه بات أحد الموضوعات الرئيسة التي تدرس تحت علم اللغة التطبيقي؛ وقد أسهمت مصادر البيانات النصية الحديثة – أعني المدونات اللغوية المحوسبة وخصوصاً مدونات المتعلمين (Learner Corpora)، مع ما يوفره الحاسب من أدوات للبحث والتحليل – في فتح الباب أمام الباحثين لدراسات أوسع في هذا المجال، حيث تسهم دراسات تحليل الأخطاء في عدة مجالات حاسوبية ولغوية إلى جانب المجال الرئيس المتمثل في تعليم اللغة الثانية وتعلمها، مما يساعد على بناء كتب دراسية، ومواد تعليمية، ومعاجم طلابية، ذات قدرة أكبر على الوفاء بأهداف العملية التعليمية، وباحتياجات المتعلمين أنفسهم؛ ومن هنا ينطلق موضوع هذا البحث الذي يهدف إلى ربط تحليل الأخطاء بالتطورات الحديثة في مجال المصادر اللغوية، شارحاً منهجية تحليل الأخطاء، ومستنداً على الأسس النظرية التي يقوم عليها هذا الميدان، مع تطبيق عملي على نماذج مستقاة من اللغة الطبيعية لمتعلمي اللغة العربية.

٤ . ٢ . المقصود بتحليل الأخطاء

يعرف لودلينق وهيرشمان (Lüdeling & Hirschmann 2015) تحليل الأخطاء Error Analysis (EA) بأنه أداة من الأدوات البحثية، وأن من لوازم استخدام هذه الأداة النظر في بعض النقاط، مثل تحديد الأخطاء المدروسة وتصنيفها، وشمول السياق اللغوي بالتحليل؛ أما أليس وباركويزن (Ellis & Barkhuizen 2005) فيعرفان تحليل الأخطاء عملياً بأنه دراسة الأخطاء اللغوية التي يقع فيها متعلمو اللغة أثناء أحاديثهم أو كتاباتهم، ويتضمن ذلك مجموعة من الإجراءات للتعرف على أخطاء المتعلمين ووصفها وتفسيرها؛ ويمكن نظرياً أن تحدث هذه الأخطاء أثناء عملية الاستقبال اللغوي (Comprehension) أو أثناء عملية الإنتاج (Production)، ولكن لصعوبة اكتشاف الأخطاء المتعلقة بعملية الاستقبال بسبب استحالة تحديد موقع الخطأ بدقة، فإن تحليل الأخطاء يركز على تلك التي تحدث أثناء عملية الإنتاج.



٤ . ٣. الفرق بين الخطأ Error والغلط Mistake

لسنا هنا بصدد دراسة الفرق اللغوي بينهما[١]، لكننا سنركز على الفرق فيما يخص مجال تعلم اللغة واكتسابها؛ حيث يفرق كوردر (Corder 1967) بين أخطاء الأداء (Performance) التي تقع من المتعلم أثناء الحديث أو الكتابة، وبين أخطاء الكفاية (Competence) التي مرجعها إلى خلفيته المعرفية عن اللغة التي يتعلمها، ويقتبس كوردر Corder تسمية هذين النوعين من ميلر (Miller 1966) الذي يسمي النوع الأول «Mistake» (غلط) والثاني «Error» (خطأ)؛ كما أن هناك نوعاً ثالثاً للأخطاء أشار إليه صيني (١٤٠٢هـ) نقلاً عن كوردر (Corder 1973) في كتابه Introducing Applied Linguistics وهو زلة اللسان (Lapes)، وفرَّق بين هذه الأنواع الثلاثة بأن زلة اللسان (Lapes) تعني الأخطاء الناجمة عن تردد المتكلم وما شابه ذلك، والغلط (Mistake) هو ذلك الناتج عن إتيان المتكلم بكلام لا يناسب الموقف، وأما الخطأ (Error) فهو ذلك النوع من الأخطاء التي يخالف فيها المتحدث أو الكاتب قواعد اللغة.

وبالرجوع إلى تقسيم ميلر (Miller 1966)، فهو كما ذكرنا يصنف النوع الأول «الغلط» تحت أخطاء الأداء اللغوي (Language performance)، وهي في الغالب أخطاء غير منتظمة، ولا تعكس أي دور في عملية تعلم اللغة؛ أما النوع الثاني «الخطأ» ففي الغالب يقع بسبب نقص في الكفاية اللغوية لدى المتعلم، ويمكن تصنيف هذه الأخطاء تحت أخطاء الكفاية المرحلية (Transitional competence)، وهي في غالبها أخطاء منتظمة، وتعكس جزءاً من النظام المعرفي اللغوي لدى المتعلم؛ ومع هذا التصنيف لأخطاء المتعلمين فإن كوردر Corder يؤكد على أن التفريق بين الغلط (Mistake) والخطأ (Error) واحدة من الصعوبات التي تواجه الباحث، فالحكم على أخطاء المتعلم يحتاج إلى دراسة وتحليل دقيقين، ولا يكفي رصدها فقط، ويوافقه في تصنيف الأخطاء وصعوبة الكشف عن نوعها لودلينق وهيرشمان (Lüdeling & Hirschmann 2015) اللذان يشيران إلى أن التفريق بين هذين النوعين ممكن نظرياً، لكنه قد يكون صعباً عملياً؛

١- يميز أبو هلال العسكري في «الفروق اللغوية» (١٩٩٧م: ٥٥) بين الخطأ والغلط بما يلي: «أن الغلط هو وضع الشيء في غير موضعه، ويجوز أن يكون صواباً في نفسه، والخطأ لا يكون صواباً على وجه». وجاء في «لسان العرب» لابن منظور (١٤١٤هـ) مادة (غلط): «الغَلَطُ: أَن تَعْيا بالشيء فلا تَعْرِفَ وجه الصواب فيه»، وفي مادة (خطأ): «الخَطَأُ والخَطاءُ: ضدُّ الصواب».



لأنه لا سبيل لإدراك ما في ذهن المتعلم من المعرفة اللغوية من ناحية – مما قد يكون سبباً في أخطاء الكفاية المرحلية – كما أنه لا سبيل لمعرفة الظروف المحيطة التي قد تؤثر على أداء المتعلم من ناحية أخرى – والتي قد تكون السبب في أخطاء الأداء اللغوي–؛ وحتى عند دراسة حالات كثيرة لخطأ واحد وقع من ذات المتعلم فقد يصعب اكتشاف ما إذا كان بسبب نقص في المعرفة اللغوية، أو أنه مجرد خطأ مطبعي وقع فيه الطالب بشكل متكرر لمشكلة لديه في استخدام لوحة المفاتيح مثلاً.

٤ . ٤ . أهمية تحليل الأخطاء

يرى كوردر (Corder, 1967) – وهو من أوائل من كتبوا في مجال تحليل الأخطاء – أن الأخطاء اللغوية قادرة على أن تكشف لنا بعض الجوانب المتعلقة بالنظام اللغوي الذي يستخدمه المتعلم في مرحلة زمنية محددة أثناء تعلم اللغة، ويؤكد على أن المتعلم لا يخرج عن كونه يستخدم نظاماً لغوياً وإن لم يكن هذا النظام صحيحاً بعد. كما يرى أن أهمية الأخطاء اللغوية تكمن في ثلاثة جوانب:

الأول: للمعلم، حيث تخبره الأخطاء عند تحليلها بطريقة منهجية عن مدى تقدم المتعلم نحو الهدف اللغوي المرسوم له، وبالتالي كم بقي لتحقيق هذا الهدف.

الثاني: للباحث، حيث تساعده الأخطاء اللغوية في الكشف عن آلية تعلم اللغة واكتسابها، وما هي الإستراتيجيات والإجراءات التي يوظفها المتعلم خلال فترة التعلم.

الثالث: لمتعلم اللغة نفسه، وقد نعتبره أهم هذه الجوانب، حيث إنه لا غنى للمتعلم عن هذه الأخطاء التي هي بمثابة أداة يستخدمها لاختبار فرضياته حول طبيعة اللغة التي يتعلمها، ومن ثم يحصل على تغذية راجعة حول هذه الأخطاء، مما يساعده على بناء نظامه اللغوي، ويمكن اعتبار الأخطاء اللغوية إستراتيجية يوظفها المتعلم عند تعلمه للغة الثانية، وحتى الطفل عند اكتسابه لغته الأم، وقد مثل كوردر (Corder, 1967) لذلك بحوار بسيط بين أم وطفلها، ظهر فيه اختبار الطفل لبعض الجمل الخاطئة حتى وصل للاستخدام الصحيح بمساعدة والدته.

وترى قرانجر (Granger, 2002) أن تحليل الأخطاء غالباً ما يثير ردات فعل رافضة لمثل هذا النوع من الدراسات بسبب عودتها إلى تلك الأزمنة السابقة التي كان ينظر فيها إلى الأخطاء اللغوية على أنها مجرد عيوب في لغة المتعلم، وتؤكد قرانجر على أن الحقيقة هي عكس ذلك، فتحليل الأخطاء مفتاح مهم في الدراسات التي تسهم في



فهم التطور الحاصل في اللغة المرحلية للمتعلم مما يجعلها عملية أساسية في بناء الأطر المنهجية والتربوية لتعليم اللغة، فالمعلمون ومصممو المواد التعليمية يحتاجون إلى معلومات دقيقة للجوانب اللغوية التي يُتوقع من الطالب أن يكون قادراً على اكتسابها في كل مرحلة، وذلك لتقديم الدخل اللغوي الأنسب لهذه المراحل، ومن هنا كان تحليل الأخطاء مصدراً قيماً لمثل لهذه المعلومات، ولا بد هنا من التنويه على أن ما سبق لا يعني تركيز التدريبات والأنشطة الصفية على الأخطاء، لكن وعي المعلم بطلابه سيقود حتماً إلى نتائج ذات جودة أعلى.

٤ . ٥ . نبذة تاريخية عن تحليل الأخطاء

لو حاولنا الرجوع إلى بداية الاهتمام بالأخطاء اللغوية في لغتنا العربية، فسنجد من الروايات-مع ضعفها-ما يشير إلى اهتمام النبي ﷺ بتقويم اللسان عند وجود اللحن[١]، أما الآثار التي وردت عن الصحابة الكرام - رضوان الله عليهم - فهي كثيرة وقد بسطها الباحثون في الكتب التي تحدثت عن نشأة النحو العربي، وقد لا يكون من المناسب إعادة بسطها هنا، لكن هذه الروايات تشير إلى أن تفشي الأخطاء كانت السبب في نشأة علم النحو، وبذلك يمكننا القول إن الاهتمام بالأخطاء قد سبق وضع قواعد النحو العربي.

ويشير الباحثان أليس وباركويزن (Ellis & Barkhuizen 2005) إلى أن تاريخ تحليل الأخطاء - فيما يخص اللغة الإنجليزية - يعود إلى نحاة القرن الثامن عشر الذين اهتموا بما يصح استخدامه لغوياً وما لا يصح، فكانت منهجية تحليل الأخطاء لديهم مبنية على هذا الأساس الذي يشبه «قُل ولا تَقُل»، وقد استمر ذلك إلى القرن الماضي إذ بدأت تظهر كتب ومعاجم تتحدث عن الأخطاء الشائعة لمتعلمي اللغة الثانية أو الأجنبية، ومنها على سبيل المثال «الأخطاء الشائعة في اللغة الإنجليزية» Fitikides (1936) Common Mistakes in English؛ أما الدراسات التي ركزت على البحث في آلية تعلم واكتساب اللغة باستخدام تحليل الأخطاء فلم تبدأ إلا في ستينيات القرن الماضي عندما حلت دراسات تحليل الأخطاء محل الدراسات التقابلية (Contrastive

١- روى الحاكم في مستدركه (٢/ ٤٣٩) عن أبي الدرداء رضي الله عنه قال: «سمع النبي ﷺ رجلا قرأ فلحن فقال رسول الله ﷺ أرشدوا أخاكم»، ضعفه الألباني في سلسلة الأحاديث الضعيفة (٢/ ٣١٥) رقم (٩١٤).



Analysis)، ولم تكن الدراسات التقابلية تمثل منهجيةً لتحليل الأخطاء، بل كان الهدف منها إجراء مقارنة بين نظامين لغويين يمثلان في الغالب لغتين مختلفتين، الأولى هي اللغة الأم للمتعلم والثانية هي لغته الهدف، ويتضمن ذلك وصف السمات اللغوية لكل من اللغتين محل الدراسة مع بيان أوجه الشبه والاختلاف بينهما، ومن ثم التنبؤ بالأخطاء التي يمكن أن يقع فيها متعلمو اللغة الهدف، وكانت هذه الدراسات تُبنى على الاعتقاد بأن الأخطاء اللغوية تحدث في الغالب نتيجة التداخل الخاطئ بين قواعد اللغة الأم وقواعد اللغة الهدف، ومن أمثلة هذه الدراسات التقابلية دراسة كفنر (Kufner 1962) بعنوان «التراكيب النحوية للإنجليزية والألمانية» The Grammatical Structures of English and German، وكذلك دراسة ستوكويل وآخرون (Stockwell et al. 1965) بعنوان «التراكيب النحوية للإنجليزية والإسبانية» The Grammatical Structures of English and Spanish؛ ومن الانتقادات التي وجهت لهذه الدراسات أن كثيراً من الأخطاء التي تنبأت بها لم تقع، وأن بعض الأخطاء التي تقع من المتعلمين لم تتنبأ بها؛ وبناء على هذه النتائج العملية – إضافة إلى رفض الأسس النظرية التي تقوم عليها الدراسات التقابلية والتي من أهمها النظرية السلوكية (انظر تشومسكي Chomsky 1957) – فقد حاول الباحثون إيجاد منهج بديل لدراسة اكتساب اللغة الثانية، فكان أول توجههم إلى منهج تحليل الأخطاء، وممن كتب عن هذا المنهج في نهاية الستينيات ومطلع السبعينيات باحث يدعى كوردر Corder –طُبعت مقالاته بعد ذلك في (١٩٨١)– حيث فَصَّل الحديث حول الأسس النظرية والإجراءات العملية لمنهج تحليل الأخطاء، وقد بات هذا المنهج يمثل الأداة الرئيسة لإجراء الأبحاث حول تعلم اللغة الثانية واكتسابها، لكن تربعه على المناهج الأخرى لم يدم طويلاً، ففي منتصف السبعينيات بدأت مناهج أخرى في الظهور ومنها على سبيل المثال «تحليل الحالات المفترضة»[1] Obligatory Occasions)

١– يهدف هذا المنهج إلى تحليل سمات لغوية محددة في اللغة الهدف للمتعلم، بقياس دقة استخدامه لهذه السمات مقارنة بما يفترض أن يكون عليه السياق الصحيح، وذلك للكشف عن مدى اكتسابه لهذه السمات، ويمكن استخدام الصيغة المختصرة التالية لاستخراج دقة استخدام السمة اللغوية محل الدراسة، ومتابعة نموها لدى المتعلم:

$$\text{دقة الاستخدام (نسبة مئوية)} = \frac{\text{عدد المرات الصحيحة لاستخدام السمة اللغوية}}{\text{عدد المرات المفترضة لاستخدام السمة اللغوية}}$$



Analysis) الذي كتب عنه براون (Brown 1973)، ثم طورته بيكا (Pica 1983) بعد ذلك وأطلقت عليه «تحليل الاستخدام الشبيه بالهدف»[1] (Target-Like Use Analysis)، وهذا لا يعني انتهاء منهج تحليل الأخطاء، فقد طوَّره بعض الباحثين وابتكروا وسائل جديدة لاستخدامه في فهم آلية تعلم اللغة واكتسابها ومنهم تايلر (Taylor, 1988) الذي اقترح منهجية نوعية (qualitative) جديدة لدراسة الأخطاء تستند إلى ما يسميه بالتقاليد التفسيرية لقواعد السلوك الإنساني، ومن الباحثين أيضاً باردوفي-هارلق وبافمن (Bardovi-Harlig & Bofman 1989) اللتين استخدمتا تحليل الأخطاء لقياس الدقة اللغوية بإجراء مقارنة بين أخطاء مجموعتين من المتعلمين تختلفان في المستوى اللغوي، كذلك استعرض لينن (Lennon 1991) طريقة للتغلب على المشاكل التي قد تطرأ عند تحديد الأخطاء وذلك بالنظر إلى السياق الأوسع للخطأ، كما قدم أليس وباركويزن (Ellis & Barkhuizen 2005) منهجية من خمس مراحل لتحليل الأخطاء، وسنأخذ مثالاً عملياً عليها في ثنايا هذا البحث؛ هذه الدراسات وغيرها أثبتت أن منهج تحليل الأخطاء لا يزال مناسباً ومستخدماً إلى اليوم للبحث في آلية تعلم اللغة واكتسابها.

ولو اتجهنا صوب الدراسات العربية التي بحثت أخطاء متعلمي اللغة العربية الناطقين بغيرها فسنجد أن ظهورها كان حديثاً، ومنها دراسة العصيلي «الأخطاء الشائعة في الكلام لدى طلاب اللغة العربية الناطقين بلغات أخرى» (١٤٠٥هـ)، ودراسة العتيق «تحليل الأخطاء الدلالية لدى دارسي اللغة العربية من غير الناطقين بها في مادة التعبير الكتابي» (١٤١٢هـ)، ودراسة الحمد «تحليل أخطاء التعبير الكتابي لدى المستوى المتقدم من دارسي العربية غير الناطقين بها في جامعة الملك سعود» (١٤١٤هـ)، ودراسة العقيلي «تحليل الأخطاء في بعض أنماط الجملة الفعلية للغة العربية في الأداء الكتابي لدى دارسي المستوى المتقدم» (١٤١٥هـ)، وغيرها من الدراسات.

ما يهمنا أن نعرفه عن طريق هذه النبذة التاريخية هو أن الغرض الرئيس من دراسة

١- هو نفس منهج « تحليل الحالات المفترضة» مع احتساب حالات الاستخدام غير المفترضة، أي التي يستخدم فيها المتعلم السمة اللغوية في غير محلها، ويعبر عنها بالتعميم الزائد (Over generalisations)، وصيغته كما يلي:

$$\text{دقة الاستخدام (نسبة مئوية)} = \frac{\text{عدد المرات الصحيحة لاستخدام السمة اللغوية}}{\text{عدد المرات المفترضة لاستخدام السمة اللغوية + عدد حالات التعميم الزائد}}$$



الأخطاء اللغوية رغم تاريخها الممتد كان يُركز على بيان الاستخدام اللغوي الصحيح وتمييزه عن الاستخدام الخاطئ، كما أن هذه الدراسات لم تكن تعزل أخطاء المتعلمين الأجانب عن أخطاء الناطقين الأصليين لتدرسها في سياق تعلم اللغة، ويتبع ذلك عدم الاهتمام بالبحث في كيفية تعلم واكتساب اللغة، أو بناء النظام المعرفي اللغوي لدى المتعلم، واكتشاف إستراتيجات التعلم لديه، ونحو ذلك من الأغراض والدراسات التي لم توجد إلا في وقتنا الحاضر.

٤ . ٦ . الأسس النظرية لتحليل الأخطاء

كما ذكرنا في النبذة التاريخية أعلاه، فإن منهج الدراسات التقابلية–الذي سبق منهج تحليل الأخطاء–قد اعتمد في أساسه على النظرية السلوكية التي ترى أن تعلم اللغة يشبه إلى حد كبير عملية ميكانيكية لتشكيل عادة سلوكية جديدة، وأن العادات التي سبق تشكيلها تتداخل مع العادات الجديدة فتسبب تشويشاً لها؛ ولذلك يكمن التحدي الذي يواجه متعلم اللغة الثانية، وكذلك المعلم، في التغلب على ما قد يقع من تأثير للعادة التي سبق تشكيلها (اللغة الأولى للمتعلم) في العادة الجديدة (اللغة الثانية)، فأخذ منهج الدراسات التقابلية على عاتقه مهمة تحديد أوجه الاختلاف بين لغة المتعلم الأم ولغته الثانية، وذلك لتسهيل اكتساب العادة الجديدة (اللغة الثانية) عن طريق الممارسة المكثفة لأوجه الاختلاف بين اللغتين. وفي مقابل منهج الدراسات التقابلية، برز منهج تحليل الأخطاء الذي ارتبط ارتباطاً وثيقاً بالاتجاهات الفطرية في تعلم اللغة، ففي حين تؤكد النظرية السلوكية على دور المثيرات البيئية، فإن النظريات الفطرية تركز على العمليات الذهنية التي تحدث في الدماغ عند تعلم اللغة، فالبيانات اللغوية (المدخلات) تُعالَج داخلياً بواسطة قدرة إدراكية مستقلة تُنتج نظاماً معرفياً لغوياً يُستخدم في الأداء اللغوي الفعلي (المخرجات)، هذا النظام المعرفي اللغوي للمتعلم يطلق عليه اللغة المرحلية (Interlanguage)، وأول من أطلق هذا المصطلح هو سيلينكر (Selinker 1972) للإشارة إلى النظام العقلي لقواعد اللغة الثانية في ذهن الدارس في مرحلة محددة من عملية التعلم، ورغم التطور الكبير الذي مرت به نظرية اللغة المرحلية خلال العقود الماضية إلا أن أساساتها ظلت ثابتة كما هي، ومنها:

١ . أن اللغة المرحلية للمتعلم تتكون من نظام معرفي لغوي ضمني، مع تعذر



الوصول لحقيقة القواعد التي تحكم هذه اللغة المرحلية.

٢. اللغة المرحلية تُمثل نظاماً لغوياً للمتعلم كما أن للمتحدث الناطق بلغته الأم نظامه اللغوي، وهذا قد يفسر الاتساق الذي يظهر في لغة المتعلم أثناء استخدامه اللغة الثانية.

٣. اللغة المرحلية للمتعلم هشـّة، فهي غير مكتملة وغير مستقرة، ومن السهل خرقها عن طريق ما يتلقاه المتعلم من الدخل اللغوي الخارجي، ومن العمليات الداخلية مثل التعميم الزائد (Over generalisations).

٤. اللغة المرحلية للمتعلم نظام انتقالي مرحلي كما هو واضح من تسميتها، فالمتعلم يبني قواعد لغته مع الوقت، وهذا يعني انتقال المتعلم عبر سلسلة من المراحل.

٥. اللغة المرحلية قابلة للتغير، فالمتعلم قد يغير التراكيب والقواعد التي يستخدمها خلال أي مرحلة من مراحل تطور نظامه المعرفي اللغوي، ورغم أن جزءاً يسيراً من هذا التغير قد يكون عشوائياً إلا أن أغلبه محكوم وفق هذا النظام الذي يمثل اللغة المرحلية.

٦. اللغة المرحلية نتاجٌ للإستراتيجيات العامة للتعلم، إذ تؤثر فيها وفق نسبة توظيفها، ومنها تطبيق قواعد اللغة الأم على اللغة الهدف (L1 Transfer)، والتعميم الزائد (Over generalisations)، والتبسيط (Simplification)، وغيرها.

٧. قد يلجأ المتعلم إلى إتمام لغته المرحلية ببعض الإستراتيجات التواصلية مثل إعادة الصياغة (Paraphrasing) ونحوها، وذلك لتعويض النقص المعرفي لديه عن اللغة الهدف، أو للتغلب على الصعوبات التي يجدها عند الحديث بها.

٨. اللغة المرحلية قد تتحجَّر، أي يتوقف نموها عند المتعلم، وبالتالي لا تصل إلى مستوى المتحدث الأصلي [(١)].

ومما يجدر الإشارة إليه أن بعض هذه الأساسات لا زالت محل نقاش، فالباحثون مثلاً غير مجمعين على أن اللغة المرحلية قابلة للتغير، ومنهم من يفضل التعامل مع التغير الذي يطرأ عليها على أنه مظهر من مظاهر الأداء اللغوي (Performance) لا الكفاية اللغوية (Competence)، انظر قريق على سبيل المثال (Gregg 1990)؛ ومن ذلك أيضاً نظرة الباحثين إلى إستراتيجات التعلم المسؤولة عن تطور اللغة المرحلية التي لا زالت محل نقاش، فمنهم

١- للاستزادة حول التحجر اللغوي لمتعلمي اللغة العربية، يمكن الاطلاع على البحث التالي: العصيلي، عبدالعزيز (١٤٢٦هـ)، التحجر في لغة متعلمي اللغة العربية الناطقين بغيرها، مجلة جامعة أم القرى لعلوم الشريعة واللغة العربية وآدابها، ١٧ (٣٣): ٣٠١-٣٨٩.



من يرى أنها لغوية بحتة، ومنهم من يرى أنها تتضمن عمليات أوسع ذات طابع إدراكي، ؛ وقد أسهم تحليل الأخطاء في تأييد(Ellis 2002) وأليس ،(White 1989) انظر مثلاً وايت كثير من هذه الأسس، إذ وجد الباحثون أن الأخطاء التي يرتكبها المتعلمون ذات نظام محدد وليست عشوائية، وأن هذه الأخطاء تتغير بمرور الوقت مما يشير إلى ما يحدث لنظام اللغة المرحلية من تطور، إضافة إلى ما يزودنا به تحليل الأخطاء من أدلة حول التحجر اللغوي، وتتجلى الصلة الوثيقة بين تحليل الأخطاء ونظرية اللغة المرحلية عن طريق مقالات «التي طبعت في كتاب بعنوان «تحليل الأخطاء واللغة المرحلية (Corder 1981) كوردر (Ellis & Barkhuizen, 2005) Error Analysis and Interlanguage.

٤ . ٧. منهجية تحليل الأخطاء

تشير قرانجر (Granger 2002) إلى أن تحليل الأخطاء في الوقت الحالي يختلف عنه في سبعينات القرن الماضي، إذ أن الأخطاء في ذلك الوقت كانت تُحلل بعيداً عن سياقاتها اللغوية، ودون النظر إلى الجوانب الصحيحة في لغة المتعلم، ولم تكن توجد تقسيمات موحدة ومعيارية لتصنيف الأخطاء اللغوية، أما تحليل الأخطاء اليوم فإنه يراعي السياق بشكليه: السياق اللغوي للخطأ (co-text)، وسياق الاستخدام (context)، وكلاهما متاح اليوم في كثير من مدونات المتعلمين، إضافة إلى وجود بعض التقسيمات الموثقة والمشروحة لتصنيف الأخطاء اللغوية.

ويَستخدم الباحثون في الغالب أدوات حاسوبية تساعدهم على البحث والتحليل عند إجراء دراساتهم في مجال تحليل الأخطاء، كما تقدم هذه الأدوات وظائف أخرى مفيدة مثل إحصاء المفردات، واستخراج قوائم الشيوع للمفردات والتراكيب (Frequency lists)، واستخراج المتصاحبات اللغوية (Collocations)، واسترجاع السياقات (Concordances)، والمقارنة بين النصوص وغيرها، ومن الأدوات المناسبة للعمل على النصوص أو المدونات اللغوية العربية: برنامج غواص (Althubaity and Almazrua 2014; Althubaity et al. 2014)، وموقع (Sketch Engine Kilgarrif 2014; Kilgarriff et al. 2004)، وبرنامج(AconCorde Roberts, 2014; Roberts et al. 2006)، ويمكن الرجوع لدراسة الفيفي وأتويل (Alfaifi



& Atwell 2016) وفيها تقويم لمجموعة من أدوات تحليل المدونات اللغوية العربية.

يقدم أليس وباركويزن (Ellis & Barkhuizen 2005) طريقة تشمل خمس مراحل لدراسة الأخطاء مستقاة مما كتبه كوردر (Corder, 1974) في أبحاثه، وهي باختصار:

- جمع عينة من لغة المتعلمين.
- تحديد الأخطاء اللغوية في العينة.
- وصف الأخطاء.
- تفسير الأخطاء.
- تقويم الأخطاء.

وفيما يلي تفاصيل أكثر حول الإجراءات والإشكاليات المرتبطة بكل مرحلة من هذه المراحل الخمس، مأخوذة مما كتبه أليس وباركويزن (Ellis & Barkhuizen 2005)، مع تطبيقات على نماذج من اللغة الطبيعية لمتعلمي اللغة العربية.

٤ . ٧ . ١ . جمع عينة من لغة المتعلمين

من المهم عند جمع العينة أن تكون مواصفاتها مناسبة لسؤال البحث الذي نحاول الإجابة عنه، إذ أن اختيار العينة قد يؤثر على نوعية الأخطاء الموجودة فيها، فأخطاء الكتابة تختلف عن أخطاء الكلام، وأخطاء الأطفال تختلف عن أخطاء البالغين، وأخطاء الكلام العفوي المباشر قد تختلف عن أخطاء الكلام المعد سلفاً، وأخطاء المتعلمين في المستوى المبتدئ قد تختلف عن أخطاء المستوى المتقدم؛ ولذلك نحتاج إلى مراعاة مجموعة من العوامل المؤثرة المتعلقة بالطالب وبالمهمة اللغوية التي يُستكتب أو يُستنطق عن طريقها، وذلك مثل العمر، والجنس، واللغة الأم، والمستوى اللغوي، ونوع النص (مكتوب أو منطوق)، والنوع الأدبي (سردي، أو حجاجي، أو مناقشة، أو غير ذلك)، وهل المهمة محدودة بوقت لإنهائها، وهل سيعطى وقتاً للإعداد قبل المهمة، وهل سيسمح للمتعلم باستخدام مراجع أثناء الكتابة أو الكلام، ونحو ذلك من العوامل.

ويمكن التمييز بين طريقتين لجمع العينة، الأولى أن يركز الباحث على عينة محددة وفق العوامل المذكورة أعلاه، فيجمع مثلاً نصوصاً منطوقة سردية من طلاب بالغين ذكور في المستوى المتقدم ولغتهم الأم هي الفرنسية، بحيث يطلب منهم الحديث مباشرة دون إعداد مسبق، ثم تليها مهمة مشابهة ولكن مع إعطائهم مهلة كافية للإعداد، ومثل



هذه العينة المركَّزة تساعد على الإجابة عن أسئلة بحثية محددة مثل: ما أثر الإعداد المسبق للمهام الشفهية السردية على أخطاء متعلمي اللغة الفرنسيين الذكور البالغين في المستوى المتقدم؟ أما الطريقة الثانية فيجمع فيها الباحث مجموعة عامة من البيانات تشمل مثلاً المكتوب والمنطوق، من عدة فئات عمرية، ومستويات لغوية، ذكوراً وإناثاً، وهكذا؛ وهنا يحتاج الباحث لوصف مكونات هذه العينة وصفاً دقيقاً ليتمكن لاحقاً من دراسة العوامل المؤثرة على أخطاء المتعلمين.

وبغض النظر عن الطريقة التي يستخدمها الباحث في جمع بياناته، فإنه قد يجمع عينته بنفسه، أو يعتمد على مدونة لغوية جاهزة إن وجد فيها ما يطابق مواصفات دراسته وأسئلته البحثية، والاعتماد على مدونة لغوية – مصممة وفق معايير واضحة – بات ممارسة شائعة في الأبحاث اللغوية النظرية والتطبيقية على حد سواء، فهي تجعل عينة البحث محايدة، ونتائجه قابلة للتعميم؛ فيعمد الباحث إلى اختيار جزء محدد من هذه المدونة إن كان سيبحث وفق الطريقة الأولى (المركَّزة)، أو يبحث فيها كاملة أو في عينة عشوائية إن كان سيعتمد على الطريقة الثانية (العامة).

وفي بحثنا الحالي سنعمد إلى استخدام عينات من نصوص المدونة اللغوية لمتعلمي اللغة العربية (Alfaifi 2015)، وتضم هذه المدونة اللغوية مجموعة من المواد المكتوبة والمنطوقة التي حررها متعلمو اللغة العربية في المملكة العربية السعودية بين عامي ٢٠١٢م و٢٠١٣م، وتشمل ١٥٨٥ نصاً (٢٨٢,٧٣٢ كلمة)، شارك في تحريرها ٩٤٢ طالباً من ٦٧ جنسية، و٦٦ لغة أم مختلفة؛ فأما المواد التي تشكل محتوى المدونة فمنها نصوص مكتوبة يدوياً، وأخرى باستخدام الحاسب، وكذلك مقابلات صوتية مسجلة، تدور جميعها حول موضوعين: الأول سردي يقص فيه الطالب رحلة في إحدى الإجازات، والثاني يناقش فيه الطالب ميوله الدراسي وتخصصه في المستقبل؛ وأما الطلاب المشاركون فهم من متعلمي اللغة العربية الناطقون بغيرها ويمثلون ٤٧٪ من محتوى المدونة، إضافة إلى طلاب ناطقين بالعربية يمثلون ٥٣٪، وهم من كلا الجنسين (الذكور ٦٧٪، والإناث ٣٥٪)، يدرسون في عدد من المدارس الثانوية ومعاهد اللغة وجامعات المملكة العربية السعودية، وتتفاوت أعمارهم بين ١٦ و ٤٢ سنة؛ ويمكن تنزيل نصوص المدونة في ملفات نصية (txt)، أو ملفات لغة الترميز الممتدة (xml)،



إضافة إلى الأصول المكتوبة يدوياً في ملفات (pdf)، وكذلك التسجيلات الصوتية في ملفات (mp3)، عن طريق موقع المدونة على الشبكة العنكبوتية على الرابط التالي: http://www.arabiclearnercorpus.com.

٤. ٧. ٢. تحديد الأخطاء اللغوية في العينة

من المهم الاتفاق على أساس أو تعريف للخطأ لينطلق منه الباحث في حكمه على ما هو خطأ وما ليس بخطأ في لغة المتعلم، وقد يكون هذا الأساس أو المعيار مستنداً على «قواعد اللغة»، أي أن ما خالف القواعد اللغوية يعد خطأ، وقد يكون المعيار قائماً على «الاستخدام اللغوي»، أي أن ما خالف الاستخدام المتداول بين الناطقين باللغة فهو خطأ لغوي، وهذا المعيار الثاني قد يكون من حيث المنهج العلمي بعيداً بعض الشيء عن الموضوعية المطلوبة في مثل هذه الحالة؛ لأنه يخضع لرأي الباحث وتقويمه بالدرجة الأولى، مما يعني احتمال وجود تفاوت كبير بين الباحثين في تحديد ما هو خطأ وما هو صواب بناء على «الاستخدام اللغوي»، كما أن استخدام هذا المعيار يميل إلى الاعتماد على الأسلوب اللغوي الذي يصعب وضع ضابط دقيق له، أكثر من اعتماده على القواعد اللغوية المتفق عليها بدرجة كبيرة، إضافة إلى أهمية معرفة سياق الاستخدام (context) للحكم على الخطأ وفق هذا المعيار، فما قد يكون خطأ في سياق قد يكون صواباً في غيره؛ وكل هذه الأسباب تجعل معيار «الاستخدام اللغوي» أقل موثوقية من معيار «قواعد اللغة» (Ellis & Barkhuizen 2005)؛ وبناء على ما سبق فقد يكون من الصعب تعريف الخطأ بدرجة مقبولة من الموضوعية بناء على معيار «الاستخدام اللغوي»، وكان لينن (Lennon 1991: 182) ممن حاول تعريفه وفق هذا المعيار فقال إن الخطأ هو: «الصيغة اللغوية -أو مجموعة الصيغ اللغوية- التي في غالب الظن لا يستخدمها الناطق الأصلي باللغة في نفس السياق الذي وردت فيه، وفي نفس الظروف التي استخدمت فيها»، وقد يسبب هذا التعريف كما ذكرنا إشكالاً منهجياً من الناحية العملية، فإذا كان الباحث نفسه هو الذي سيحكم على الصيغ اللغوية -أي: ما إذا كان الناطق باللغة يستخدمها أو لا يستخدمها- فليس هناك ما يضمن موضوعية الحكم، وقد يكون الحكم على الخطأ من عدة أطراف أحد الحلول المقبولة لهذا الإشكال؛ أما تحديد الخطأ وفق معيار «قواعد اللغة» فيمكننا أن نقول مثلاً بأن الخطأ هو ما خالف



قواعد اللغة المتفق عليها فيما يتعلق بالأصوات أو المفردات أو التراكيب، سواء أكان ذلك من حيث الرسم أم الصرف أم النحو أم الدلالة، أو نحواً من ذلك.

وبغض النظر عن المعيار الذي نستخدمه لتحديد الخطأ، فإننا نحتاج إلى القيام ببعض الخطوات العملية عند تحديد الأخطاء، ويمكن تلخيصها فيما يأتي:

أولا: إعادة بناء العينة-سواء أكانت نصاً مكتوباً أم حديثاً منطوقاً-لتبدو وكأن ناطقاً باللغة هو الذي أنتجها.

ثانياً: مقارنة الجمل في النصين-نص المتعلم مع النص الذي أعاد بناءَه الناطق باللغة-وعند وجود اختلافات بينهما ننظر في إمكانية قبول الجملة التي أنتجها المتعلم وموافقتها لمعيار الخطأ الذي نتبناه (القواعد أو الاستخدام)، وتبقى بعد ذلك الجمل التي لا يمكن قبولها فنكون بذلك قد حصرنا الجمل المشتملة على أخطاء.

ثالثاً: تضييق مجال الخطأ أكثر، وذلك بتحديد الكلمة أو الكلمات الخاطئة في كل جملة من الجمل غير المقبولة في نص المتعلم، لنصل بذلك إلى تحديد دقيق للأخطاء اللغوية في العينة التي بين أيدينا.

لكن إشكالاً قد يواجهنا في الخطوة الأولى، وهو صعوبة الوصول إلى الصيغة الأصلية عند إعادة بناء العينة، فقد يحتمل البناء الصحيح أكثر من شكل، وكل واحد من هذه الأشكال يشير إلى سبب مختلف للخطأ، وتبقى الصيغة النهائية التي يقصدها المتعلم غير معروفة، ومثال ذلك عبارة «أن كل حاج رجال الذين أحرموا بالحج ليسوا لباس الإحرام»[1]، فمثل هذه الجملة قد تحتمل التصويبات التالية:

أن كل حاج رجل من الذين أحرموا بالحج لبسوا لباس الإحرام.

أن كل الحجاج الرجال الذين أحرموا بالحج لبسوا لباس الإحرام.

وقد تحتمل تصويبات أخرى، لكننا سنلاحظ أن كل تصويب يشير إلى مجموعة من المشاكل منها ما هو مشترك بين التصويبين ومنها ما ليس كذلك، فالتصويب الأول يشير إلى أن المشكلة تكمن في:

١- يمكن الاطلاع على السياق الكامل للجملة في النص رقم S821_T1_F_Pre_NNAS_W_H في المدونة اللغوية لمتعلمي اللغة العربية (www.arabiclearnercorpus.com).



- المطابقة في العدد [رجال-رجل].
- سقوط حرف الجر [من].
- الخطأ الإملائي في إبدال الباء ياءً في الفعل [لبسوا].

والتصويب الثاني يشير إلى أن المشكلة تكمن في:

- التعريف والتنكير [حاج-الحجاج] و[رجال-الرجال].
- المطابقة في العدد [حاج-الحجاج].
- الخطأ الإملائي في الفعل [لبسوا].

ويرى كوردر (Corder 1974) أن أحد الحلول الممكنة لمثل هذه المسألة أن يُسأل المتعلم عن مقصده من حديثه أو نصه، ليسهل الوصول إلى أقرب صيغة تمثل هذا المقصد، لكن هذا الحل ممكن فقط في حال استطاع الباحث معرفة المتعلم صاحب النص وأمكنه الوصول إليه، ومع ذلك فإن جيمس (James 1998) يرى أن الأخطاء في الغالب تبقى غامضة بالنسبة للمتعلم، فيستحيل عليه تبعاً أن يحدد أي الصيغ هي الأنسب لما يقصده بكلامه أو كتابته.

وفي بعض الحالات يساعدنا السياق على استنتاج الصيغة المقصودة، فالعبارة التالية مثلاً: «وسألنا بعض السؤال من الفقه والحديث وغبر ذالك»[1] لا يتضح فيها ما إذا كان المتحدث سائلاً أو مسؤولاً، فقد يكون المقصود منها:

وسألناه بعض الأسئلة من الفقه والحديث وغير ذلك (المتحدث وزملاؤه هم السائلون).

وسألونا بعض الأسئلة من الفقه والحديث وغير ذلك (المتحدث وزملاؤه هم المسؤولون).

لكن الجملة التي بعدها «وأعطوا مكافأة الذين إستطعوا الجواب سؤالهم» تدل على أن المتحدث وزملاءه كانوا مسؤولين لا سائلين، لأن من يجيب على السؤال يُعطى

١- هذه الجملة من النص رقم S515_T1_M_Pre_NNAS_W_C في المدونة اللغوية لمتعلمي اللغة العربية.



مكافأة، فيكون التصويب الثاني هو الأقرب لمقصود المتعلم، وبالتالي نجد أن الخطأ يتركز في (١) إسقاط الضمير (الواو) [سألنا-سألونا] أو في رسم الهمزة إن كان الفعل مبنياً للمجهول [سألنا-سُئِلنا]، (٢) المطابقة في العدد [السؤال-الأسئلة]، (٣) أخطاء الرسم الإملائي [غبر-غير] و[ذالك-ذلك].

٤ . ٧ . ٣ . وصف الأخطاء

وصف الأخطاء كما يشرحه كوردر (Corder 1974: 128) «هو في الأساس عملية مقارنة بين النص الأصلي المشتمل على الأخطاء والنص الصحيح الذي أعيد بناؤه»، فهذه العملية تتضمن تحديد الاختلافات بين الصيغ التي يستخدمها المتعلم وتلك التي قد يستخدمها الناطق باللغة، وبالتالي فالوصف يُركِّز على السمات الظاهرة في نص المتعلم، وهذه المرحلة - وصف الأخطاء - تشمل خطوتين:

الأولى: وضعُ نظام لتصنيف الأخطاء ووصفها، يُستخدم في وسم الأخطاء بعد تحديدها في النص.

الثانية: إحصاء مدى شيوع كل خطأ من هذه الأخطاء.

ويشير جيمس (James 1998) إلى اثنين من المعايير الواجب توفرها عند بناء نظام لتصنيف الأخطاء ووصفها، الأول أن يحتوي هذا التصنيف على تفاصيل تسمح له بوصف الأخطاء بأعلى درجة من الدقة، والثاني أن يكون واضحاً للمستخدم فلا يشتمل على أخطاء غامضة وغير مفهومة، وقد أشار أليس وباركويزن (Ellis & Barkhuizen 2005) إلى مجموعة من التصنيفات المستخدمة لوسم الأخطاء في اللغة الإنجليزية، أما فيما يخص اللغة العربية فهناك عدة دراسات تحدثت عن تحليل الأخطاء التي يقع فيها متعلمو اللغة، ورغم أن هذه الدراسات قد صنَّفتْ بعض أنواع الخطأ معتمدة على عينات واقعية مأخوذة من المتعلمين (انظر مثلاً العصيلي، ١٤٠٥هـ، والعتيق، ١٤١٢هـ، والحمد، ١٤١٤هـ، والعقيلي، ١٤١٥هـ)، إلا أنها لم تضع تصنيفاً شاملاً للأخطاء يمكن اعتباره مرجعاً في وسم المدونات اللغوية، وقد يرجع السبب إلى أن مدونات المتعلمين لم تكن قد اكتسبت شهرتها كنوع مستقل من المدونات، إذ كانت السبب الرئيس في ظهور مثل هذه التصنيفات العامة للأخطاء اللغوية.



ويُعدّ تصنيف الأخطاء الذي قدمه أبو حكيمة وآخرون (Abuhakema et al. 2008, 2009) من المحاولات الأولى في هذا المجال، وهو في الأصل ترجمة لتصنيف الأخطاء في اللغة الفرنسية الذي أوردته قرانجر (Granger 2003) في دراسة لها، انظر الفيفي وأتويل (Alfaifi & Atwell 2012) لمزيد من الإيضاح حول هذا التصنيف؛ وهناك تصنيف آخر بناه الباحث (Alfaifi et al. 2013) وهو جزء من مشروع المدونة اللغوية لمتعلمي اللغة العربية (Alfaifi 2015)، وفيه طبقتان، الأولى تشمل مجالات الخطأ، وقد رُتبتْ تبعاً لتسلسل المستويات اللغوية كما يلي: الخطأ الإملائي، الخطأ الصرفي، الخطأ النحوي، الخطأ الدلالي، الخطأ في علامات الترقيم، وتحت كل مجال من هذه المجالات الخمسة مجموعة من الفئات الفرعية (انظر الجدول ١)؛ وقد أسس الباحث تصنيف الأخطاء هذا مستنداً على نتائج مجموعة من الأبحاث العلمية التي درست عينات من أخطاء متعلمي اللغة (انظر الفيفي وأتويل Alfaifi & Atwell 2012)، ومن ثم اختبر التصنيف على عينات من المدونة اللغوية لمتعلمي اللغة العربية بمشاركة مجموعة من الباحثين، والذين شاركوا أيضاً في تقويم هذا التصنيف عدة مرات مستقلة لتحسينه وتنقيحه حتى وصل إلى نسخته الثالثة، وفي كل مرة يُطبَّق على عينة مختلفة من النصوص للتأكد من دقته وشموله ووضوح أنواع الخطأ فيه؛ كما درس الباحث استخدام هذا التصنيف مقارنة بتصنيف أبي حكيمة – وقد اشترك الباحثان في هذه الدراسة (الفيفي وأبو حكيمة) –، وأظهرت النتائج أن استخدام تصنيف الفيفي حقق نتائج أعلى خصوصاً فيما يتعلق بتحديد المجال العام للخطأ، وللاطلاع على تفاصيل أكثر عن هذه الدراسة يُنظَر الفيفي وآخرون (Alfaifi et al. 2013).

ويعرض الجدول (١) أدناه فئات الخطأ في تصنيف الفيفي مع الأنواع الفرعية المندرجة تحت كل فئة، وقد أنشأ الباحث دليلاً مرجعياً يشرح فيه كل نوع من أنواع الخطأ في هذا التصنيف مع أمثلة للتوضيح، كما يشرح الطريقة المناسبة لاختيار الرمز الصحيح لكل خطأ، ويزود المستخدم ببعض النقاط والقواعد الهامة التي يجب اتباعها عند عملية وسم الأخطاء، ويشرح عدداً من حالات التداخل المحتملة وطريقة التعامل معها[1]، وهذا الدليل جزء من مشروع المدونة اللغوية لمتعلمي اللغة العربية التي سبق

١– يمكن الاطلاع على دليل وسم الأخطاء عن طريق الرابط التالي:
http://www.abdullahalfaifi.com/Error_Tagging_Manual_v2.pdf.



ذكرها، هذا فيما يخص الخطوة الأولى ضمن مرحلة وصف الأخطاء، وهي «وضعُ نظام لتصنيف الأخطاء ووصفها، يُستخدم في وسم الأخطاء بعد تحديدها في النص».

مجال الخطأ	نوع الخطأ	الرمز
الإملاء	١. Hamza (ء، أ، إ، ؤ، ئ، ئـ) الخطأ في الهمزة	<OH>
	٢. Confusion in Hā’ and Tā’ Mutadarrifatīn الخطأ في الهاء والتاء المتطرفتين (ـه، ـة، ـت)	<OT>
	٣. Confusion in ‘alif and Yā’ Mutadarrifatīn (ا، ى، ي) الخطأ في الألف والياء المتطرفتين	<OA>
	٤. Confusion in ’alif Fāriqa (كتبوا) الخطأ في الألف الفارقة	<OW>
	٥. Confusion between Nūn (ن) and Tanwīn («ـً ـٍ») الخلط بين النون والتنوين	<ON>
	٦. Shortening the long vowels تقصير الصوائت الطويلة (اوي «ـُـِ»)	<OS>
	٧. Lengthening the short vowels تطويل الصوائت القصيرة (اوي «ـُـِ»)	<OG>
	٨. Wrong order of word characters الخطأ في ترتيب الحروف داخل الكلمة	<OC>
	٩. (Replacement in word characters) استبدال حرف أو أحرف من الكلمة	<OR>
	١٠. (Redundant characters) زيادة حرف أو أكثر	<OD>
	١١. (Missing characters)s نقص حرف أو أكثر	<OM>
	١٢. Other orthographical errors أخطاء إملائية أخرى	<OO>



مجال الخطأ	نوع الخطأ	الرمز
الصرف	١٣. Word inflection الخطأ في اختيار بنية الكلمة المناسبة	<MI>
	١٤. Verb tense الخطأ في زمن الفعل	<MT>
	١٥. Other morphological errors أخطاء صرفية أخرى	<MO>
النحو	١٦. Case الخطأ في الإعراب	<XC>
	١٧. Definiteness الخطأ في التعريف والتنكير	<XF>
	١٨. Gender الخطأ في الجنس (التذكير والتأنيث)	<XG>
	١٩. (Number singular, dual and plural) الخطأ في العدد (الإفراد والتثنية والجمع)	<XN>
	٢٠. Redundant word كلمة زائدة	<XT>
	٢١. Missing word كلمة ناقصة	<XM>
	٢٢. Other syntactic errors أخطاء نحوية أخرى	<XO>
الدلالة	٢٣. Word selection الخطأ في اختيار الكلمة المناسبة	<SW>
	٢٤. Fasl wa wasl (confusion in use/non-use of conjunctions) الخطأ في الفصل والوصل (الخطأ في استخدام أدوات العطف)	<SF>
	٢٥. Other semantic errors أخطاء دلالية أخرى	<SO>



مجال الخطأ	نوع الخطأ	الرمز
علامات الترقيم	٢٦. Punctuation confusionعلامة ترقيم خاطئة	<PC>
	٢٧. Redundant punctuation علامة ترقيم زائدة	<PT>
	٢٨. Missing punctuation علامة ترقيم مفقودة	<PM>
	٢٩. Other errors in punctuation أخطاء أخرى في علامات الترقيم	<PO>

الجدول (١): النسخة الثالثة من تصنيف الأخطاء اللغوية العربية الذي طوره الباحث.

أما الخطوة الثانية وهي «إحصاء مدى شيوع كل خطأ من هذه الأخطاء»، فالمثال التالي يشرحها، إذْ أخذ الباحثُ عينة من المدونة اللغوية لمتعلمي اللغة العربية تشمل ١٠,٠٠٠ كلمة، تمثل الناطقين بالعربية والناطقين بغيرها (٥٠٠٠ كلمة لكل فئة)، ووَصَفَ ثلاثةُ مختصين الأخطاءَ في هذه العينة عن طريق وضع الوسم المناسب أمام كل خطأ (الوسم <OH> مثلاً أمام أي خطأ في الهمزة)، وأظهرت النتائج أن ترك علامات الترقيم، والخطأ في الهمزة، هما الأكثر شيوعاً بين متعلمي اللغة العربية، ويوضح الجدول ٢ أكثر عشرة أخطاء شيوعاً في العينة الموسومة.

م	مجال الخطأ	نوع الخطأ	النسبة*
١	علامات الترقيم	علامة ترقيم مفقودة	٪٢٣
٢	الإملاء	الخطأ في الهمزة	٪١٩
٣	الدلالة	الخطأ في اختيار الكلمة المناسبة	٪٧
٤	علامات الترقيم	علامة ترقيم خاطئة	٪٧
٥	النحو	كلمة زائدة	٪٥



م	مجال الخطأ	نوع الخطأ	النسبة*
٦	النحو	الخطأ في التعريف والتنكير	٥٪
٧	الإملاء	الخطأ في الهاء والتاء المتطرفتين (ـه، ـة، ـت)	٥٪
٨	النحو	كلمة ناقصة	٥٪
٩	الدلالة	الخطأ في الفصل والوصل (الخطأ في استخدام أدوات العطف)	٤٪
١٠	الإملاء	نقص حرف أو أكثر	٣٪
		المجموع	٨٣٪

الجدول (٢): الأخطاء الأكثر شيوعاً في عينة المدونة اللغوية لمتعلمي اللغة العربية.[١]

ولو وزعنا الأخطاء بناء على عامل أصالة اللغة (Nativeness)، فسنجد أن أكثر الأخطاء شيوعاً لدى الناطقين بالعربية هما: الخطأ في الهمزة، يليه ترك علامة الترقيم؛ أما بالنسبة للناطقين بغير العربية فإن أكثر الأخطاء شيوعاً لديهم على الترتيب: ترك علامة الترقيم، والخطأ في التعريف والتنكير، والخطأ في اختيار الكلمة المناسبة، ووجود كلمة زائدة؛ ويوضح الجدول (٣) توزيع الأخطاء على الناطقين بالعربية والناطقين بغيرها، ويمكن الرجوع إلى دراسة الفيفي (Alfaifi 2015) للاطلاع على تفاصيل أكثر حول عمليتي الوسم والإحصاء.

م	الناطقون بالعربية			الناطقون بغير العربية		
	مجال الخطأ	نوع الخطأ	النسبة	مجال الخطأ	نوع الخطأ	النسبة
١	الإملاء	الخطأ في الهمزة	٢٨٪	علامات الترقيم	علامة ترقيم مفقودة	١٨٪

١ – ملاحظة: نسبة الخطأ إلى كامل العينة الموسومة.

م	الناطقون بالعربية			الناطقون بغير العربية		
	مجال الخطأ	نوع الخطأ	النسبة	مجال الخطأ	نوع الخطأ	النسبة
٢	علامات الترقيم	علامة ترقيم مفقودة	٪٢٦	النحو	الخطأ في التعريف والتنكير	٪١٢
٣	الإملاء	الخطأ في الهاء والتاء المتطرفتين	٪٧	الدلالة	الخطأ في اختيار الكلمة المناسبة	٪١١
٤	علامات الترقيم	علامة ترقيم خاطئة	٪٦	النحو	كلمة زائدة	٪١٠
٥	الدلالة	الخطأ في اختيار الكلمة المناسبة	٪٥	النحو	كلمة ناقصة	٪٨
٦	النحو	الخطأ في الإعراب	٪٤	علامات الترقيم	علامة ترقيم خاطئة	٪٨
٧	الدلالة	الخطأ في الفصل والوصل	٪٤	النحو	الخطأ في الجنس (التذكير والتأنيث)	٪٥
٨	الإملاء	نقص حرف أو أكثر	٪٣	الدلالة	الخطأ في الفصل والوصل	٪٤
٩	الإملاء	استبدال حرف أو أحرف من الكلمة	٪٣	الإملاء	الخطأ في الهمزة	٪٤
١٠	النحو	كلمة زائدة	٪٣	الصرف	الخطأ في اختيار بنية الكلمة المناسبة	٪٤
	المجموع		٪٨٨	المجموع		٪٨٤

الجدول (٣): الأخطاء الأكثر شيوعاً موزعة على الناطقين بالعربية والناطقين بغيرها.



ولنأخذ مثالاً للبحث عن إحدى الكلمات؛ للكشف عن دقة استخدامها من قبل الناطقين بغير العربية، فلو بحثنا في كامل المدونة عن استخدام كلمة «يَتَعَلَّق» لدى الناطقين بغير العربية، سنجد أنها استخدمت ٢٨ مرة، ويمكن أن نستبعد واحدة منها؛ لأن الكلمة نفسها تكررت في ذات الجملة ولم ترد بينهما مفردةٌ صحيحة، وبالتالي فإن لدينا ٢٧ حالة لاستخدام كلمة «يتعلق»، أتى حرف الجر «الباء» بعد خمس عشرة حالة منها تمثل ٥٦٪، بينما أتت كلمات أخرى متنوعة بعد الحالات الباقية التي تمثل ٤٤٪ من مجموع الحالات (انظر هذه الحالات في الجدول ٤)، وبما أن كلمة «يتعلق» يعقبها في الغالب حرف الجر «الباء» نحو: «تعلق بكذا» إن لم يفصل بينهما بالفاعل نحو «تعلق فلان بكذا» فيمكن تصنيف هذا الخطأ على أنه من الأخطاء الدلالية، وهو «Word selection الخطأ في اختيار الكلمة المناسبة» في عشر حالات وهي التي استخدم فيها المتكلم أو الكاتب كلمة لا ترد بعد الفعل «يتعلق» بناء على ما نجده من أمثلة لها في المعاجم العربية[1] (الأمثلة من ١ إلى ١٠ في الجدول ٤)، وأما المثالان الباقيان فيمكن تصنيف الخطأ فيهما على أنه «Missing word كلمة ناقصة»، إذ أسقط الكاتب حرف الباء من السياق دون استبدال حرف آخر به (الأمثلة ١١ و١٢ في الجدول ٤).

م	السياق السابق	مفردة البحث مع الكلمة التالية	السياق اللاحق	نوع الكلمة بعد مفردة البحث	وصف الخطأ
١	لدكاتر يوضح الناس ما	يتعلق في	ذالك ثم بعد	حرف جر	الخطأ في اختيار الكلمة المناسبة
٢	من يوضح الناس ما	يتعلق في	ذالك، فسئلت في	حرف جر	الخطأ في اختيار الكلمة المناسبة
٣	التخصص لأن هذا التخصص	يتعلق في	جيمع أحكام الله	حرف جر	الخطأ في اختيار الكلمة المناسبة

١- في المعجم الوسيط مثلاً: (تَعَلَّقَ) الشوك بالثوب: عَلِق، و-الوحشُ أو الظبيُ بالحِبالة: وقع فيها وأمسكته. و-الإبلُ: أكلت العَلْقَى. و -الشيءَ: عَلَّقه. و-فلاناً، وبه: أحبَّه (المعجم الوسيط، ١٤٢٦هـ).



م	السياق السابق	مفردة البحث مع الكلمة التالية	السياق اللاحق	نوع الكلمة بعد مفردة البحث	وصف الخطأ
٤	أعرف أكثر من أمور	يتعلق في	الدين والشريعة كما	حرف جر	الخطأ في اختيار الكلمة المناسبة
٥	الطلاب في الفصل عما	يتعلق عن	النصيحة وشروط الرحلة	حرف جر	الخطأ في اختيار الكلمة المناسبة
٦	لهم، ثم ألقيت محاضرة يتعلق أ	يتعلق عن	الفوائد	حرف جر	الخطأ في اختيار الكلمة المناسبة
٧	كثيرا من المعلومات التي	يتعلق على	الأحاديث والتفسير،	حرف جر	الخطأ في اختيار الكلمة المناسبة
٨	لي أبي قصةً ما	يتعلق على	حياتي، أخبرني في	حرف جر	الخطأ في اختيار الكلمة المناسبة
٩	الإخوة سألقي إليكم محاضرة	يتعلق أن	طلبة العلم طلب	حرف ناسخ	الخطأ في اختيار الكلمة المناسبة
١٠	مثل هذا التخصص ومما	يتعلق حول	هذا التخصص لي	ظرف	الخطأ في اختيار الكلمة المناسبة
١١	فإنه نشعر بالارتياح وفيما	يتعلق التخصص	أو المجال الذي	اسم	كلمة ناقصة
١٢	اللغة، وميراثنا العلمي فيما	يتعلق الشرعية	وغيرها لا يمكن	اسم	كلمة ناقصة

الجدول (٤): الكلمات الواردة بعد الفعل «يتعلق» من غير مصاحباتها الصحيحة.



وعند وسم الأخطاء في هذه الأمثلة فإن الوسم سيكون للكلمة الواردة بعد الفعل في الأمثلة العشرة الأولى (لأنها غير مناسبة للسياق)، وأما الاثنان الأخيران فسيكون للفعل نفسه لوجود كلمة مفقودة بعده مباشرة (وهي حرف الجر الباء)، كما في الجدول ٥.

م	المثال
١	لدكاتر يوضح الناس ما يتعلق في <SW> ذالك ثم بعد
٢	من يوضح الناس ما يتعلق في <SW> ذالك، فسئلت فى
٣	التخصص لأن هذا التخصص يتعلق في <SW> جيمع أحكام الله
٤	أعرف أكثر من أمور يتعلق في <SW> الدين والشريعة كما
٥	الطلاب في الفصل عما يتعلق عن <SW> النصيحة وشروط الرحلة
٦	لهم، ثم ألقيت محاضرة يتعلق أ يتعلق عن <SW> الفوائد
٧	كثيرا من المعلومات التي يتعلق على < SW> الأحاديث والتفسير ،
٨	لي أبي قصةً ما يتعلق على <SW> حياتي، أخبرني في
٩	الإخوة سألقي إليكم محاضرة يتعلق أن <SW> طلبة العلم طلب
١٠	مثل هذا التخصص ومما يتعلق <SW> حول هذا التخصص لي
١١	فإنه نشعر بالارتياح وفيما يتعلق <XM> التخصص أو المجال الذي
١٢	اللغة، وميراثنا العلمي فيما يتعلق < XM> الشرعية وغيرها لا يمكن

الجدول (٥): وسم حالات الخطأ الواردة مع الفعل «يتعلق».

٤ . ٧ . ٤ . تفسير الأخطاء

يرى الباحثون في اكتساب اللغة الثانية أن هذه المرحلة هي أهم مراحل تحليل الأخطاء، إذ تتضمن تحديد مصدر كل خطأ للوقوف على أسباب حدوثه، والذي يعنينا



هنا هو المصدر اللغوي النفسي للخطأ، أي العمليات الذهنية لاستخدام اللغة الثانية وطبيعة نظامها المعرفي في ذهن المتعلم، ويرى رامبتن (Rampton 1987) وجود مصادر لغوية اجتماعية للأخطاء، ومنها أن يعمد المتعلم إلى استخدام صيغة خاطئة – مع معرفته بخطئها – كوسيلة للتحكم في الانطباع الذي يريد إيصاله للطرف الآخر، ومن ذلك مثلاً التلطف في الرفض باستخدام أساليب توحي بأنه لا يزال متعلماً بسيطاً للغة.

أحد الأسباب البارزة للوقوع في الأخطاء يرجع إلى الصعوبة التي يواجهها المتعلم عند استرجاع المعلومات اللغوية المتعلقة باللغة الثانية أثناء التواصل المباشر مع الآخرين، فالصيغ والقواعد التي لم يسيطر عليها المتعلم بعد، تتطلب منه جهداً ذهنياً كبيراً، ولذا قد يلجأ إلى استخدام صيغ وقواعد سبق له استخدامها والسيطرة عليها مع أنها خاطئة.

وعند تفسير الأخطاء نحتاج إلى معرفة العمليات الذهنية التي قد يستدعيها المتعلم عند عدم إلمامه بالصيغ أو القواعد الصحيحة للغة الثانية، وهنا لا بد من التفريق بين نوعين أساسيين من العمليات يؤدي كل نوع منها إلى فئة مختلفة من الأخطاء: أخطاء سببها التداخل بين اللغات (interlingual)، وأخطاء نابعة من داخل اللغة الثانية نفسها (intralingual).

فالأخطاء التي يسببها التداخل بين اللغات تنتج في الغالب من تأثير قواعد وصيغ لغة أخرى – لغته الأم أو أي لغة يعرفها – على اللغة الثانية، ويذكر كوردر (Corder, 1983) عمليتين تحت هذا النوع من الأخطاء: النقل (Transfer) والاقتراض (Borrowing)، ويحدث النقل عند إقحام المتعلم لقواعد لغة أخرى في نظامه اللغوي المعرفي المسمى باللغة المرحلية لتصبح جزءاً منها، أما الاقتراض فهو استخدام مؤقت لصيغ وقواعد لغة أخرى كإستراتيجية للتواصل فقط، ولا يستلزم إدراج هذه الصيغ والقواعد في نظام اللغة المرحلية بشكل دائم؛ وليس من السهل تحديد ما إذا كانت إحدى الصيغ أو القواعد المستخدمة في اللغة الهدف هي نتيجة لعملية النقل من لغة أخرى، ولكي نستطيع الجزم بأن النقل هو سبب الخطأ فإننا نحتاج إلى إجراء مقارنة بين متعلمين من خلفيات لغوية مختلفة، لنكتشف أن هذا الخطأ يصدر عن مجموعة طلاب يشتركون في لغة واحدة، ولا يقع فيه أولئك الذين ينتمون إلى اللغات الأخرى.

أما الأخطاء النابعة من داخل اللغة الثانية نفسها فهي انعكاس لإستراتيجيات التعلم



العامة، أي التي تظهر لدى كل المتعلمين بغض النظر عن لغاتهم الأم، ويستعرض جيمس (James 1998) قائمة بأهم هذه الإستراتيجيات:

١. القياس الخاطئ (False analogy): وهو نوع من التعميم الزائد (Over generalisations) الذي سبق ذكره، كاستخدام الجمع السالم لما يجمع جمع تكسير (حديقة يجمعها على حديقات بدل حدائق).

٢. التحليل الخاطئ (Misanalysis): ومن ذلك اعتقاد أن الاسم «حمزة» مؤنث بسبب التاء المربوطة، أو أن «عثمان» مثنى بسبب الألف والنون في آخره.

٣. استخدام قواعد ناقصة (Incomplete rule application): ويعبر عنه بالتعميم الناقص (Under generalisations)، ومن ذلك الخطأ في موقع اسم الاستفهام «أين» في جملة «وضعتَ الكتاب أين؟» قياساً على ما يحل محلها في جملة الجواب «وضعتُ الكتاب على الطاولة».

٤. استغلال الزيادة (Exploiting redundancy): مثل إسقاط الأجزاء التي لها وظيفة نحوية لكنها لا تؤدي إلى تغيير كبير في المعنى، مثل حذف علامة الجمع [الواو والنون] في «رأيت الطلاب وهم يذهب إلى المدرسة».

٥. تجاهل قيود التلازم اللغوي (Overlooking co-occurrence restrictions): فمع أن لفظتي «عاجلة» و«سريعة» مترادفتان، إلا أن «وجبات عاجلة» ليست من المتلازمات اللفظية بخلاف «وجبات سريعة».

٦. التبسيط (Simplification): أي اللجوء إلى تبسيط القواعد التي قد تحتاج إلى جهد لتعلمها، مثل استخدام صيغة المفرد في جميع الحالات «أَعرفُ كلَّ الطلاب الذي نجحوا في الاختبار».

ومن المشاكل المرتبطة باستخدام هذه القائمة أو غيرها أنه قد يكون من الصعب تحديد أي هذه الاستراتيجيات هي المسؤولة عن وقوع الخطأ، فالخطأ الواحد قد تنطبق عليه أكثر من إستراتيجية.

ويمكن النظر إلى الخطأ بطريقة أخرى، فيمكن أن يكون خطأ طَبَعياً (natural)، أي ناتجاً عن إستراتيجيات التعلم العامة التي سبق استعراضها، أو خطأ مُحدَثاً (induced)، ويُعرِّف ستينسن (Stenson 1974) الخطأ المحدَث بأنه ذلك الخطأ الناتج عن طريقة تعليم اللغة، ويمثل له بالخطأ أثناء التدريبات الشفهية في القاعة الدراسية، ويمكننا

تنزيله على المثال التالي:

المعلم: يا خالد، اسأل زميلك أحمد: أين ذهب في الإجازة الماضية؟

خالد: يا أحمد، أين ذهب في الإجازة الماضية؟

لكن أليس وباركويزن (Ellis & Barkhuizen 2005) يريان أن هذا الخطأ يمكن أن يكون طبعياً ويمكن أن يكون مُحدثاً في نفس الوقت، فإذا أرجعناه إلى إحدى إستراتيجيات التعلم العامة مثل استخدام قواعد ناقصة، كان بذلك طبعياً، وإن نظرنا إلى أثر عملية التعلم وتسببها في الخطأ كان خطأ مُحدثاً؛ وبشكل عام فإن تحديد مصدر الخطأ ليس من الأمور السهلة، وذلك لكثرة المصادر التي لم يُتعرَّف عليها بدقة إلى الآن، وحتى لو حصل ذلك فإن الإشكالات المرتبطة بهذه المصادر قد تستمر في الظهور، إذ أن الخطأ نفسه قد يعطينا فقط طرف الخيط الذي يوصلنا إلى مصدره، ويبقى مصدر كثير من الأخطاء غامضاً، ولهذا السبب فإن شاكتر وسيلس-مرسيا (Schachter & Celce-Murcia 1977) يوصيان الباحثين بالحذر الشديد عند التعرف على مصدر أحد الأخطاء، إذ أن كثيراً من الأخطاء قابلة للتفسير بناء على مصادر متعددة أكثر من ارتباطها بمصدر وحيد.

وبالرجوع لقائمة الأخطاء الواردة في الجدول ٤ والتي تَسرِد مجموعة من حالات الخطأ في ما يَرِدُ بعد الفعل «يتعلق»، فإن أقرب ما يمكن أن تُفسَّر به الأخطاء العشرة الأولى هو «تجاهل قيود التلازم اللغوي Overlooking co-occurrence restrictions»، إذ أن حرف الجر «الباء» من ملازمات الفعل «يتعلق»، وقد استبدَلَ بها المتكلم أو الكاتب كلمة أخرى، مما جعل في الكلام خللاً واضحاً، وقد يُفسَّر الخطأ في المثالين الأخيرين (١١ و١٢)، حيث سقط حرف الجر الباء، بأنه من قبيل «استغلال الزيادة Exploiting redundancy»، أي أن يُسقط الكاتب الأجزاء التي لها وظيفة نحوية لكنها لا تؤدي إلى تغيير كبير في المعنى.

ومثل هذه التفسيرات لا يمكن القطع بها كما ذكرنا، فقد نجد لدى المتكلم أو الكاتب - لو سألناه - تفسيراً آخر مخالفاً لما نعتقد، كما أن بعض هذه الأخطاء قد يكون من قبيل التداخل بين اللغات (interlingual) لا من داخل اللغة الثانية نفسها (intralingual)، أو من قبيل الغلط لا الخطأ، أي أنه مرتبط بالأداء اللغوي لا بالكفاية اللغوية، وبالتالي



فإننا نحتاج أولاً إلى أدلة أكثر وضوحاً لإثبات أن هذه الأخطاء تعكس النظام المعرفي اللغوي لدى المتعلم، ثم أدلة أخرى ترشدنا لمصدر هذا الخطأ، وبالتالي لتفسير أسبابه بدقة أكبر؛ ليكون البحث فيها بعد ذلك ذا نتيجة عملية.

٤. ٧. ٥. تقويم الأخطاء

تقويم الأخطاء ليس مرحلة أساسية في تحليلها بقدر ما هو إجراء إضافي للاستفادة من نتائج التحليل، وهو يتضمن تحديد وزن كل خطأ ومدى الحاجة إلى أخذه في الاعتبار عند تصميم المواد التعليمية، ويشمل تقويم الأخطاء الخطوات التالية:

١. تحديد الخطأ الذي سيتم تقويمه، وقد يشمل ذلك جميع الأخطاء المستخرجة من نص المتعلم أو مجموعة مختارة منها وهو الغالب، وتُستعرض هذه الأخطاء إما في الجمل التي وردت فيها، أو في كامل السياق النصي لها.

٢. تحديد المعيار المناسب للحكم على الأخطاء، وأشهر المعايير المستخدمة هو «وزن الخطأ»، أي مدى أهميته وخطورته على لغة المتعلم، ومن المعايير أيضاً «درجة وضوح الخطأ».

٣. إعداد أداة لتقويم الخطأ، وتتضمن في الغالب قائمة بالجمل المشتملة على الأخطاء أو النص كاملاً تتبعه طريقة التقويم، ومن أشهر هذه الطرق تقويمُ الأخطاءِ باستخدام تصنيفٍ فيه عدة درجات تبدأ من «الأكثر أهمية» وتنتهي عند «الأقل أهمية»، أو باستخدام مقياس لايكرت(Likert Scale (Likert 1932 ، الذي يمكن أن يشمل الخيارات التالية: «مهم جداً»، «مهم»، «محايد»، «غير مهم»، «غير مهم على الإطلاق»، أو نحو ذلك.

٤. اختيار التقويم المناسب لكل خطأ، ومن الأفضل أن تُقوَّم بواسطة اثنين على الأقل، وكثرة المقيمين يزيد من مصداقية النتائج ومن إمكانية تعميمها.

اشتهرت الدراسات المعنية بتقويم الأخطاء في سبعينيات وثمانينيات القرن الماضي، إلا أن شهرتها صارت أقل بسبب عدم وجود مقياس دقيق لوزن الأخطاء، ومع ذلك فإن معلم اللغة قد يحتاج إلى اتخاذ قرار حول الأخطاء الأكثر أهمية لمعالجتها، وبالتالي فهو عملياً سيحتاج إلى إجراء تقويم للأخطاء بشكل أو بآخر.



٤ . ٨. الخاتمة

سلط البحثُ الضوءَ على منهج تحليل الأخطاء، فشرح المقصود به، وبَيَّنَ أهميته، وأعطى نبذة تاريخية عنه مع شرح الأسس النظرية التي قام عليها، كما ربَطَهُ بأحد المصادر اللغوية لمتعلمي اللغة العربية، فبعد الشرح النظري لخطوات هذا المنهج طُبقت هذه الخطوات عملياً على نماذج من المدونة اللغوية لمتعلمي اللغة العربية، فقدم بذلك مثالاً بسيطاً لتحليل الأخطاء القائم على مدونات المتعلمين؛ ولعل الدراسات العربية تتوالى في هذا الميدان لتصل بإذن الله إلى نتائج أكثر شمولاً وعمقاً في تفسير أخطاء متعلمي اللغة العربية، ويكون إسهامها فاعلاً في تطوير طرائق التدريس، والمواد التعليمية، والمعاجم، وفي تعليم اللغة العربية وتعلمها بشكل عام.

قائمة المراجع

أولاً: المراجع العربية

- ابن منظور، جمال الدين (١٤١٤هـ). لسان العرب. ط٣، بيروت: دار صادر.
- الحمد، محمد ماجد (١٤١٤هـ). تحليل أخطاء التعبير الكتابي لدى المستوى المتقدم من دارسي العربية غير الناطقين بها في جامعة الملك سعود. رسالة علمية غير منشورة. الرياض: جامعة الإمام محمد بن سعود الإسلامية.
- صيني، محمود إسماعيل (١٤٠٢هـ). التقابل اللغوي وتحليل الأخطاء. ط١، الرياض: عمادة شؤون المكتبات، جامعة الملك سعود.
- العتيق، زايد مهلهل (١٤١٢هـ). تحليل الأخطاء الدلالية لدى دارسي اللغة العربية من غير الناطقين بها في مادة التعبير الكتابي. رسالة علمية غير منشورة. الرياض: جامعة الإمام محمد بن سعود الإسلامية.
- العسكري، أبو هلال (١٩٩٧م). الفروق اللغوية. تحقيق محمد إبراهيم سليم. القاهرة: دار العلم والثقافة.
- العصيلي، عبدالعزيز إبراهيم (١٤٠٥هـ). الأخطاء الشائعة في الكلام لدى طلاب اللغة العربية الناطقين بلغات أخرى: دراسة وصفية تحليلية. رسالة علمية غير منشورة. الرياض: جامعة الإمام محمد بن سعود الإسلامية.



العقيلي، عبدالمحسن سالم (١٤١٥هـ). تحليل الأخطاء في بعض أنماط الجملة الفعلية للغة العربية في الأداء الكتابي لدى دارسي المستوى المتقدم. رسالة علمية غير منشورة. الرياض: جامعة الإمام محمد بن سعود الإسلامية.

المعجم الوسيط (١٤٢٦هـ). مجمع اللغة العربية. ط٤، القاهرة: مكتبة الشروق الدولية.

ثانيًا: المراجع الأجنبية

Abuhakema, G., Faraj, R., Feldman, A. and Fitzpatrick, E. (2008). Annotating an Arabic Learner Corpus for Error. In: Pproceedings of the International Conference on Language Resources and Evaluation, LREC 2008, 26 May - 1 June 2008, Marrakech, Morocco.

Abuhakema, G., Feldman, A. and Fitzpatrick, E. (2009). ARIDA: An Arabic Interlanguage Database and Its Applications: A Pilot Study. Journal of the National Council of Less Commonly Taught Languages (JNCOLCTL), 7: 161-184.

Alfaifi, A. and Atwell, E. (2012) المدونات اللغوية لمتعلمي اللغة العربية: نظامٌ لتصنيف وترميز الأخطاء اللغوية «Arabic Learner Corpora (ALC): A Taxonomy of Coding Errors». The 8th International Computing Conference in Arabic (ICCA 2012). Cairo, Egypt.

Alfaifi, A. and Atwell, E. (2016). Comparative evaluation of tools for Arabic corpora search and analysis. International Journal of Speech Technology, 19 (2): 347–357.

Alfaifi, A. (2015). Building the Arabic Learner Corpus and a System for Arabic Error Annotation. PhD thesis, University of Leeds.

Alfaifi, A., Atwell, E. and Abuhakema, G. (2013). Error Annotation of the Arabic Learner Corpus: A New Error Tagset. In: Language Processing and Knowledge in the Web, Lecture Notes in Computer Science. 25th International Conference, GSCL 2013, 25-27 September 2013. Darmstadt, Germany, Springer, 14 - 22 (9), 2013.

Al-Thubaity, A., and Al-Mazrua, M. (2014). Khawas: Arabic



Corpora Processing Tool USER GUIDE. Retrieved 6 April 2014, from: http://sourceforge.net/projects/kacst-acptool/files/?source=navbar

- Al-Thubaity, A., Khan, M., Al-Mazrua, M., and Almoussa, M. (2014). KACST Arabic Corpora Processing Tool «Khawas» [Computer Software]. Retrieved 6 April 2014, from: http://kacst-acptool.sourceforge.net/

- Bardovi-Harlig, K. and Bofman, T. (1989). Attainment of Syntactic and Morphological Accuracy by Advanced Language Learners. Studies in Second Language Acquisition, 11 (1): 17-34.

- Brown, R. (1973). A first language: the early stages. Cambridge, Massachusetts: Harvard University Press.

- Chomsky, N. (1959). Review of B. F. Skinner›s Verbal Behavior. Language, 35: 26-58.

- Corder, S. (1983). A role for the mother tongue. In Gass, S. & Selinker, L. (eds.), Language Transfer and Language Learning. Rowley, Mass: Newbury House.

- Corder, S. P. (1967). The Significance of Learners› Errors. International Review of Applied Linguistics in Language Teaching, 5: 161-170.

- Corder, S. P. (1974). Error analysis. In J. Allen and S. P. Corder (eds.), The Edinburgh Course in Applied Linguistics. Volume 3: Techniques in Applied Linguistics. Oxford: Oxford University Press.

- Corder, S. P. (1981). Error Analysis and Interlanguage. Oxford: Oxford University Press.

- Ellis, N. (2002). Frequency Effects in Language Processing: A Review with Implications for Theories of Implicit and Explicit Language Acquisition. Studies in Second Language Acquisition, 24 (2): 143-188.

- Ellis, R. and Barkhuizen, G. (2005). Analysing Learner Language. Oxford: Oxford University Press.

- Fitikides, T. (1936). Common Mistakes in English. London: Longman.



Granger, S. (2002). A bird's-eye view of learner corpus research. In: Granger, S., Hung, J. and Petch-Tyson, S. (eds.), Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching (pp. 3-33). Amsterdam: John Benjamins.

Granger, S. (2003). Error-tagged learner corpora and CALL: a promising synergy. CALICO Journal, 20(3): 465-480.

Gregg, K. (1990). The Variable Competence Model of Second Language Acquisition, and Why It Isn›t. Applied Linguistics, 11 (4): 364-83.

James, C. (1998). Errors in Language Learning and Use: Exploring Error Analysis. London: Longman.

Kilgarriff, A (2014). Sketch Engine [Computer Software]. Retrieved 6 April 2014, http://www.sketchengine.co.uk/

Kilgarriff, A., Rychly, P., Smrz, P., and Tugwell, D. (2004). The Sketch Engine. In the proceedings of the Euralex, 6-10 July 2004 (pp. 105-116). Lorient, France.

Kufner, H.L. (1962). The Grammatical Structures of English and German. Chicago: University of Chicago Press.

Lennon, P. (1991). Error: Some Problems of Definition, Identification, and Distinction. Applied Linguistics, 12 (2): 180-196.

Likert, R. (1932), A Technique for the Measurement of Attitudes. Archives of Psychology, 140: 1–55.

Lüdeling, A. and Hirschmann, H (2015). Error annotation systems. In S. Granger, G. Gilquin and F. Meunier (eds.), The Cambridge Handbook of Learner Corpus Research. Cambridge: Cambridge University Press.

Miller, G. A. (1966). Language and Psychology. In E. H. Lenneberg (ed.), New Directions in the Study of Language. Cambridge: MIT Press.

Pica, T. (1983). Methods of Morpheme Quantification: Their Effect on the Interpretation of Second Language Data. Studies in Second Language Acquisition, 6 (1): 69-78.



Rampton, B. (1987). Stylistic variability and not speaking ‹normal› English: Some post-Labovian approaches and their implications for the study of interlanguage. In R. Ellis (ed.), Second Language Acquisition in Context (pp. 47-58). London: Prentice Hall International.

Roberts, A (2014). aConCorde [Computer Software]. Retrieved 6 April 2014, http://www.andy-roberts.net/coding/aconcorde

Roberts, A., Al-Sulaiti, L. and Atwell, E. (2006). aConCorde: Towards an open-source, extendable concordancer for Arabic. Corpora, vol. 1: 39-60.

Schachter, J. and Celce-Murcia, M. (1977). Some Reservations Concerning Error Analysis. TESOL, 11: 441-451.

Selinker, L. (1972). Interlanguage. International Review of Applied Linguistics in Language Teaching, 10: 209–232.

Stenson, N. (1974). Induced Errors. In J. H. Schumann & N. Stenson (eds.), New Frontiers in Second Language Learning (pp. 54-70). Massachusetts: Newbury House.

Stockwell, R.P., J.D. Bowen and J.W. Martin (1965). The Grammatical Structures of English and Spanish. Chicago: University of Chicago Press.

Taylor, D. (1988). The Meaning and Use of the Term 'Competence' in Linguistics and Applied Linguistics. Applied Linguistics, 9 (2): 148-168.

White, L. (1989). Universal Grammar and Second Language Acquisition. Amsterdam: John Benjamins.

الفصل الخامس

د. هند بنت مطلق العتيبي

المعالجة الآلية للمدونات المتوازية واستخداماتها في تعليم اللغات وتدريب المترجمين

الملخص

يتناول هذا الفصل عرضاً موجزاً لتعريف القارئ العربي بالمدونات المتوازية، التي تتضمن مجموعة كبيرة من النصوص وترجماتها بلغة (أو بلغات) أخرى يتم تخزينها إلكترونياً. وقد راجت المدونات المتوازية في مجال اللغويات التطبيقية، واكتسبت أهمية خاصة في مجالات تعليم اللغات، ودراسات الترجمة، وتدريسها، والترجمة الآلية، واللغويات المقارنة، وعلم المصطلحات، وغيرها من مجالات البحث اللغوي النظري والتطبيقي. ويستعرض هذا الفصل استخدامات المدونات المتوازية في تعليم اللغات، وفي تدريب المترجمين ثم يتناول أشهر المدونات المتوازية التي تضم اللغة العربية موضحين الحاجة الماسة لتطوير مثل هذه المصادر اللغوية. ومن ثم نستعرض وصفاً لمشروع AEPC، وهي المدونة المتوازية (الإنجليزية / العربية) التابعة لكلية اللغات والترجمة في جامعة الملك سعود، التي تهدف إلى توفير مصدر لغوي وتعليمي مجاني يسهم في دعم الأبحاث في مجالات اللغويات والترجمة وتعليم اللغات.

الكلمات المفتاحية: المدونات المتوازية، تعليم اللغات، تدريب المترجمين، اللغويات الحاسوبية، لغويات المدونات.



٥ . ١ . المقدمة

يمكن تعريف المدونة corpus بأنها قواعد بيانات تنظيمية ضخمة نسبياً لمجموعة من النصوص الإلكترونية، التي تُجمع وتصمم وفق معايير محددة للاستخدام في البحث اللغوي (انظر ليتش Leech 2010). ويمكن تقسيم أنواع المدونات حسب التعدد اللغوي؛ فقد تكون المدونة أحادية اللغة، أي تحتوي على نصوص بلغة واحدة فقط، كما قد تحتوي على لغتين أو أكثر. ومن أنواع المدونات المدونة المتوازية Parallel Corpus التي يُعرِّفها سينكلير (Sinclair 1995: 19) بأنها «نصوص أصلية وترجماتها بلغة (أو بلغات) أخرى». كما تُعرِّفها بيكر (Baker 1995) بأنها «مجموعة نصوص كتبت باللغة الأصل (أ) وترجماتها باللغة (ب)».

وقد راجت المدونات المتوازية في مجال اللغويات التطبيقية، واكتسبت أهمية خاصة في مجالات تعليم اللغات، ودراسات الترجمة، وتدريسها، والترجمة الآلية، واللغويات المقارنة، وعلم المصطلحات، وغيرها من مجالات البحث اللغوي -نظرياً كان أو تطبيقياً-.

وبالرغم من الاهتمام المتزايد بالمدونات المتوازية في الأوساط البحثية العالمية؛ فإن اللغة العربية ما زالت تفتقر لمثل هذه المصادر اللغوية (انظر السليطي وأتول Al-Sulaiti, Atwell 2006 والعجمي Al-Ajmi 2011، والمجيول Almujaiwel 2012).

ونتناول في هذا الفصل وصفاً للمدونات المتوازية، واستخداماتها المختلفة في مجال تعليم اللغات، وتدريب المترجمين، كما نستعرض أشهر المدونات المتوازية التي تتضمن اللغة العربية، ومن ثم نتطرق إلى مشروع تصميم المدونة الإنجليزية العربية المتوازية English-Arabic Parallel Corpus EAPC التابعة لكلية اللغات والترجمة بجامعة الملك سعود.

٥ . ٢ . ما المدونة المتوازية؟

ويُعرّف سينكلير Sinclair (١٩٩٥: ١٩) المدونة المتوازية بأنها «نصوص أصلية مع ترجماتها بلغة (أو بلغات) أخرى». كما تعرفها بيكر Baker 1995 بأنها «مجموعة نصوص كتبت باللغة الأصل (أ) وترجماتها باللغة (ب)».

وقد انتشر استخدام المدونات المتوازية في الأوساط البحثية في مجال اللغويات



التطبيقية، واكتسبت أهمية خاصة في مجالات تعليم اللغات (انظر بوتلي وآخرون Botley et al. 1996)، ودراسات الترجمة وتدريسها (انظر بيكر Baker 1993، 1995 وألوهان Olohan 2004)، والترجمة الآلية (انظر لُو وآخرين Lü et al. 2007 وراوف وتشونك Rauf and Schwnek 2011 وتيان وآخرين Tian et al. 2014)، واللغويات المقارنة (انظر جوهانسون Johansson 1999)، والدراسات المعجمية وعلم المصطلحات، (انظر بوكر بيرسون Bowker Pearson 2002).

ويرى العديد من الباحثين أن المدونات المتوازية تساهم في إعطاء رؤى جديدة بخصوص اللغة الأصلية واللغة المترجمة قلما يمكن الحصول عليها من دراسة المدونات أحادية اللغة، حيث تُمكِّن الباحث من المقارنة بين اللغات، والسمات الدلالية، والثقافية، وغيرها مما يزيد معرفتنا بهذه اللغات. كما تسهم هذه المصادر -كما ذكرنا آنفاً- في العديد من التطبيقات مثل: الدراسات المعجمية، والترجمة الآلية، وتعليم اللغات، ونحو ذلك (انظر مكانري وهاردي ٢٠١٦: ٣٤-٣٩).

وسنتناول في المحطات التالية من هذا الفصل استخدام المدونات المتوازية في تعليم اللغات وفي تدريب المترجمين، ثم نُعرِّج على أشهر المدونات المتوازية باللغتين العربية والإنجليزية.

٥ . ٣. استخدام المدونات المتوازية في تعليم اللغات

تزامن الاهتمام بالمدونات عموماً مع ظهور فكرة استثمارها في مجالات تعليم اللغات، فظهرت بنهاية الثمانينات وبداية التسعينات العديد من الدراسات التي تدعم ذلك الاتجاه، ومنها جونز Johns 1986 وستيفنز Stevens 1991) وغيرهما.

ويطرح رومر (Römer 2006) منهجين لاستخدام المدونات في تعليم اللغة؛ وهما: مباشر وغير مباشر. ويعتمد المنهج المباشر على استخدام الطلاب والمعلمين المدونات لاستعراض محتويات المدونة، والبحث فيها مباشرة دون وسيط، ويطلق على هذا النوع من التعلم القائم على الشواهد Data-Driven Learning ويصفه جونز (Johns 2002) بمواجهة الطالب مباشرة بالبيانات قدر الإمكان؛ بحيث تخلق من المتعلم باحثاً لغوياً (لمزيد حول هذا المنهج التربوي انظر فلورديو Flowerdew 2015 وجونز Johns 2002).

ويعتمد المنهج غير المباشر في استخدام المدونات على الدراسات البحثية؛ حيث يركز



على البراهين التي يمكن استقاؤها من المدونات، وتأثيرها على تصميم المناهج الدراسية، واختيار المواد التعليمية التي تُقَدَّم للطالب. وتنتمي مدونة المتعلمين Learners' Corpus لهذا المنهج؛ حيث تفيد هذه المدونات في تحليل أخطاء الدارسين من خلفيات ومستويات مختلفة، ثم تُسهم في دراسات اكتساب اللغة الثانية (انظر صالح ٢٠١٢).

تُعدّ المدونات المتوازية مصدراً غنياً في تعليم وتعلم اللغة؛ إذ يشير العديد من الباحثين إلى ما يسمى بالاستخدام الفعلي للغة Real language use وهو ما توفره قواعد البيانات النصية في المدونات (انظر غودوين-جونز Godwin-Jones 2001). فقد لوحظ مَيْل مصادر وكتب تعليم اللغات بطبيعتها التوضيحية، وقيود المكان الزمان إلى استخدام الأمثلة القصيرة، في حين تعكس اللغة الطبيعية ميلاً إلى التراكيب اللغوية، واستخدام السياق والاستغناء به عن اللغة المنطوقة في المواقف الطبيعية للحياة اليومية، وهو الأمر الذي يدفع بمصممي المناهج الدراسية ومدرسي اللغة إلى إعطاء أمثلة مصطنعة والاكتفاء بها عن الأمثلة الحقيقية للغة الواقعية اليومية؛ مما يُشكِّل قصوراً في معظم المناهج المؤلفة لتعليم اللغة.

ويرى مكانري (McEnery 2006) وآخرون أن المهتمين بهذا المجال من مؤلفين وباحثين بدأوا بتدارك هذا القصور، فظهرت العديد من الدراسات والكتب التي تشجع على استخدام المدونات مصدراً للغة المستخدمة في الحياة الواقعية بدلاً من الاعتماد على اللغة المتكلفة المصطنعة (انظر العصيمي، ٢٠١٢).

وتدعم المدونات الحاسوبية -عموماً- والمدونات المتوازية -بشكل خاص- المجالات المتعلقة بتدريس اللغة حيث تساهم بشكل فعال في دراسات ونظريات اكتساب اللغة وفلسفة التدريس، كما تشكل مصدراً لاختيار المادة التي تُدرَّس، وكيف تُدرَّس، وتساعد المعلمين وواضعي المناهج الدراسية على تصميم المادة التعليمية، والتدريبات المساعدة لمتعلمي اللغة، وتتيح أيضا المجال لتطوير الوسائل التعليمية والأنشطة الصفِّية وإعداد الاختبارات اللغوية (انظر آيمير Aijmer 2009؛ وانظر أيضا: فرانكنبيرج-جراسيا وآخرين Frankenberg-Garcia et al. 2011).

ويشجع استخدام المدونات في تعليم اللغة مبادئ التعلم الفعال، ومنها الاستقلالية في التعلم (autonomy) ومبدأ التعلم الذاتي، حيث يمكن للطالب الاستعانة بهذه المصادر دون الحاجة للرجوع إلى المعلم، بل يتحول المعلم إلى مرشد وموجه في العملية



التعليمية. وتتيح أدوات تحليل المدونات مثل: المكشاف السياقي (concordancer) وقوائم التكرار وقوائم المتلازمات اللفظية (colocations) للطالب استعراض المعاني والاستخدامات المختلفة للمفردات والتراكيب والتفريق بينها بعيداً عن الحدس والتخمين، وذلك من خلال نماذج وسياقات واقعية تتيح للطالب التحليل الاستقرائي وتوليد الفرضيات واختبارها والاستنتاج والتعميم.

وفي استعراض لبعض الدراسات التي تناولت استخدام المدونات المتوازية في تعليم اللغة؛ نشير إلى بحث أجراه جون(John 2001) على أحد طلاب اللغة الألمانية بالمستوى المبتدئ، حيث طلب منه الاستعانة بالمكشاف السياقي لمدونة (ألمانية/ إنجليزية) متوازية، وقد هدفت هذه الدراسة إلى وصف الكيفية التي يتعامل بها الطالب مع المدونة المتوازية، إلى جانب الاستنتاجات التي خلص إليها عند المقارنة بين اللغتين، وخصوصًا عند مقارنة العناصر المعجمية. وقد خلص الباحث إلى الإمكانيات الهائلة التي يمكن أن تقدمها المدونات المتوازية لطلاب المستويات المبتدئة على الوجه الأخص، ممن يعتمدون غالباً اعتماداً كبيراً على المعاجم العادية، وكذلك على لغتهم الأم.

وفي دراسة أخرى قام بها تشيو هاو وآخرون (Chujo et al.2009)، تمت الاستعانة بمدونة متوازية تتضمن نصوصاً مستردة آليًا من الجرائد اليابانية، وترجمتها باللغة الانجليزية لطلاب اللغة اليابانية بالمستوى المبتدئ، حيث استخدم الطلاب المكشاف السياقي ثنائيي اللغة؛ لاستعراض بعض الخصائص النحوية في نصوص المدونة. وأظهرت الدراسة تطوراً ملحوظاً في مهارات الطلاب اللغوية والمعجمية، وكذلك أبدوا ردود فعل إيجابية تجاه استخدام المدونات في التدريس.

كما تناول سانغاوا (Sangawa 2014) في دراسة بعنوان (الطالب كمعجمي) استخدام القواميس بين متعلمي اللغة اليابانية في جامعة لوبليانا في سلوفانيا، حيث ساهم استخدام المدونات المتوازية في تعزيز الوعي بالجوانب المختلفة للمفردات اليابانية من منظور مقارن، كما عرض الباحث عدداً من الأدوات والأساليب التي يمكن تطبيقها لتطوير مهارات الطلاب المعجمية، واقترح بعض استراتيجيات تطوير مهارات المفردات عن طريق استخدام المدونات.

وفي دراسة حديثة قام بها وانغ ولي (Wong and Lee 2016) باستخدام مدونة متوازية لتعليم اللغة الكانتونية للطلاب من متحدثي اللغة الصينية، قدم للطلاب



قائمة من المفردات بلغتهم الأم، وعدداً من المعاني المحتملة لها باللغة الكانتونية المعتمدة على السياق، وعن طريق الاستعانة بالمكشاف السياقي يبحث الطلاب بشكل مستقل لاستخراج أمثلة على هذه المعاني وترجماتها. وقد خلصت هذه الدراسة إلى أن استخدام المنهج التعليمي القائم على البيانات ساهم في تحسين مهارات طلاب المستوى المبتدئ في المفردات الكانتونية، وشجعت الدراسة على الاستفادة من مثل هذه المصادر التعليمية.

ويتضح من الدراسات السابقة وغيرها الامكانيات الهائلة للمدونات عموماً، والمدونات المتوازية على وجه الخصوص، والتي يمكن استثمارها بفعالية في مناهج تعليم اللغات وتعلمها. وسنتناول في المحطة القادمة كيف يمكن الاستعانة بالمدونات في تدريب المترجمين، وتطوير مهاراتهم.

٥ . ٤ . استخدام المدونات المتوازية في تدريب المترجمين

تشير العديد من المراجع إلى أن فكرة استخدام المدونات المتوازية في الترجمة قد ظهرت في بداية التسعينات على يد عدد من الباحثين من أشهرهم: منى بيكر (Mona Baker 1993)، وجون سنكلير انظر (Sinclair 1995)، وأستون (Aston 1999)، حيث انبثق مصطلح دراسات الترجمة القائمة على لغويات المدونات في دراساتهم corpus-based translation studies وبدأ يلقى رواجاً في الأوساط البحثية. وثمة جانب نظري وتطبيقي لاستخدام المدونات المتوازية في دراسات الترجمة هنستون (Hunston 2002)، ويتضمن الجانب النظري دراسة عملية الترجمة، والكيفية التي يتم فيها التعبير عن فكرة ما بلغتين مختلفتين أو أكثر، كما يشمل ذلك مقارنة بين خصائص النص الأصلي والنص المترجم. أما الجانب التطبيقي فيتعلق بالاستخدامات المختلفة للمدونات في تطوير أدوات للترجمة بمساعدة الحاسوب، كذاكرات الترجمة ونظم إدارة المصطلحات وكذلك تدريس الترجمة وتدريب المترجمين وهو ما يعنينا هنا.

ويشير غوداك (Gouadec 2002) إلى أن تعليم المترجمين يتضمن «تدريب الطالب على تنفيذ وظائف محددة ضمن أوساط محددة بالاستعانة بأدوات ونظم محددة». وتُعدّ المدونات المتوازية إحدى هذه الأدوات التي يمكن إدراجها ضمن البرنامج التدريبي، بحيث يمكن للطالب الاستعانة بها لمساعدته في مشروع الترجمة. وقد لوحظ مؤخراً اهتمام العديد من المدربين بتطوير مدونات تخصصية، وتدريب الطلاب على مهارات تطوير مدوناتهم الخاصة بحسب مشروع الترجمة الذي يعكفون عليه (سواء كان نصاً



قانونياً، أو طبياً ...إلخ). كما يحرص المدربون على تزويد الطلاب بمهارات تتعلق بمبادئ تصميم المدونات، وأنواعها، ومعايير اختيارها، وأدوات البحث، واستعراض محتوياتها. بالإضافة إلى التعرف على برمجيات معالجة المدونات، وكيفية استثمارها في مشروع الترجمة الذي يعكفون عليه.

تُسْهِم المدونات المتوازية في تعريف الطلاب بطبيعة الاختلاف بين النص الأصلي وترجمته، وكيفية نقل المعنى، أو التعبير عن نفس المضمون بلغتين (أو أكثر)، وكذلك تأثير اللغة الأم على الترجمة. كما يمكن للطالب الاستعانة بها لاستعراض المتلازمات اللفظية، والقوالب المألوفة، والتراكيب الشائعة في كلا اللغتين (الأصل، والترجمة). كما تساهم الأمثلة المُسْتَقاة من المدونات المتوازية في التحقق من خيارات الترجمة المستمدة من مصادر أخرى، كالقواميس، ومعاجم المصطلحات، أو تلك التي يفرضها الحدس واختيار الصائب منها بحسب السياق.

ويشير بيرسون (Pearson 2003) إلى أن المدونات المتوازية تتيح للطلاب الفرصة لمقارنة تراجمهم بأعمال المترجمين المحترفين، ومن ثم التعرف على الأساليب التي يمارسها أولئك المترجمين، والاستراتيجيات الفعالة التي يتبنونها. كما تساهم المدونات المتوازية في زيادة وعي الطلاب بطبيعة النصوص التخصصية (كالنصوص القانونية أو الطبية على سبيل المثال) والسمات والمصطلحات والتراكيب اللغوية المرتبطة بتلك النصوص انظر كوبلار (Kübler 2011). وتزخر الأدبيات بالعديد من الدراسات التي تتناول الأثر الإيجابي للمدونات في تطوير مهارات المترجمين، وتشجع على استثمارها ودمجها في مناهج تدريب الطلاب، ومنها-على سبيل المثال-: بيرسون(Pearson 2003) وبيرنارديني (Bernardini 2004)، وسنستعرض فيما يلي أحدثها.

ففي دراسة أجراها فرانكنبيرغ-غارسيا وسانتوس (Frankenberg-Garcia & Santos, 2003) استخدم الباحثان المدونة المتوازية comparable corpora التي تحتوي على نصوص باللغتين (البرتغالية-الإنجليزية) في تصميم تدريبات تساعد طلاب الترجمة على التعرف على السمات التقابلية بين اللغتين (الإنجليزية والبرتغالية)، والتي قد تؤدي إلى مشاكل في الترجمة، وكيف يمكن التعامل معها. وتناول الباحثان الأثر الإيجابي لهذا المنهج في تدريب المترجمين، مُخْتَتِمين بحثهما بدعوة القائمين على وضع المناهج إلى استثمار المدونات المتوازية كمصدر تعليمي لا غنى عنه.



وقد جرى العرف على الاستعانة بالمدونات المتوازية لبحث طبيعة النص المترجم عن طريق مقارنته بالنص الأصلي، إلا أن الأوساط البحثية بدأت تشهد مؤخراً اهتماماً بالاستعانة بالمدونات المتوازية لدراسة النصوص الأصلية. ففي دراسة حديثة قام بها صالحي (Salhi 2013) باستخدام مدونة الأمم المتحدة EAPCOUNT لدراسة المعاني المتعددة لكلمة Destruction، حيث نَوَّه في خاتمة الدراسة إلى فوائد الاستعانة بالمدونات المتوازية لكشف السمات الدلالية والتداولية للنص الأصلي، تماماً كما قد تفيد لبحث تلك السمات في النص المترجم.

كما قدَّم رودريغز (Rodríguez 2014) أمثلة لعدد من التدريبات التي يمكن الاستعانة بها في تدريب المترجمين من اللغة الإسبانية إلى اللغة الإنجليزية، وتعتمد على استعراض النصوص المتوازية للغتين، وشدد على أهمية هذا المصدر في رفع مستوى الوعي لدى طلاب الترجمة بالتراكيب التقابلية، والمتلازمات اللفظية، والمصطلحات الشائعة. وأظهر الاستبيان الذي أجراه في نهاية الدراسة زيادة شعور الطلاب بالثقة في تحديد التراكيب اللغوية المتكررة، والمتلازمات اللفظية بعد استعانتهم بالمدونة المتوازية.

وفي دراسة مشابهة؛ قام يان لي Haiyan Li وتشونغ شين داي(Zhongxin Dai 2014) بالاستعانة بمدونة متوازية تحوي نصوصاً مترجمة من اللغة الإنجليزية إلى الصينية في تدريب تسعين طالباً من طلاب الترجمة في جامعة صينية. حيث قسَّم الطلاب إلى مجموعتين، يتم تدريس المجموعة الأولى بالطريقة التقليدية، في حين يتم تدريب المجموعة الثانية بتدريبات قائمة على البحث في المدونة المتوازية. وبعد إجراء الاختبارات القبلية والبعدية؛ خلصت الدراسة إلى وجود تحسن ملحوظ في مهارات المجموعة الثانية، حيث عزا الباحث هذا التحسن إلى منهج التدريس القائم على استخدام المدونة المتوازية.

كما قام هيلين وفيربليتز (Heylen and Verplaetse 2015) بدراسة استخدام المدونات المتوازية في تدريب الطلاب في مقرر الترجمة الطبية، وتحليل تأثيرها على أدائهم. وقد خلصت الدراسة إلى فائدة استخدام المدونات في تدريب الطلاب على العبارات الاصطلاحية، حيث طلب من عدد من طلاب الترجمة في مرحلة (الماجستير) بترجمة عدد من النشرات الطبية من اللغة الإنجليزية للغة الألمانية (لغتهم الأم) باستخدام



مدونة متوازية تدعم اللغتين. وبمقارنة مستوى الترجمة مع المجموعة الأخرى التي لم تستخدم المدونة؛ لوحظ فرق في المستوى لصالح المجموعة الأولى.

وتشير الدراسات السابقة وغيرها الكثير إلى أهمية استثمار المدونات المتوازية في فصول تدريب المترجمين، إلا أن هناك على ما يبدو قصوراً في الدراسات التي تبحث هذا الجانب في عالمنا العربي، وتتناول طلاب الترجمة العرب. وقد يرجع ذلك إلى أن اللغة العربية ما زالت تفتقر لمثل هذه المصادر اللغوية (انظر السليطي وأتويل Al-Sulaiti and Atwell 2006 والعجمي Al-Ajmi 2011).

وسنستعرض في المحطة القادمة أشهر المدونات التي تتضمن اللغة العربية، ثم نُعرِّج على مشروع تصميم المدونة الإنجليزية العربية المتوازية AEPC التابعة لكلية اللغات والترجمة بجامعة الملك سعود، والذي يهدف إلى المساهمة في إثراء المصادر اللغوية العربية.

٥ . ٥ . المدونات العربية المتوازية

تشهد الأوساط البحثية العالمية اهتماماً متزايداً بالمدونات عموماً، والمدونات المتوازية على وجه الخصوص، إلا أن اللغة العربية ما زالت تفتقر لمثل هذه المصادر اللغوية الغنية. ويرجع ذلك -بلا شك- في أن تصميم المدونات المتوازية ينطوي على العديد من التحديات، والتي يمكن تلخيص أهمها فيما يلي:

- صعوبة معالجة اللغة العربية مقارنة باللغة الإنجليزية مثلاً؛ حيث تترابط الحروف مع بعضها لتكوِّن ما يسمى (مجموعات حروف مترابطة)، كما تستخدم اللغة العربية أنواعاً متعددة من العناصر الخارجية، مثل: النقاط، والهمزات، والمد، وعلامات التشكيل.
- صعوبة العثور على نصوص مترجمة من مجالات متنوعة، وأوعية متعددة بشكل كافٍ لتجميع مدونة متوازنة.
- قصور البرمجيات الخاصة بمعالجة اللغة العربية، مثل برامج المحاذاة، والترميز، والتنقيح، والمحللات الصرفية، وغيرها.
- قلة الدعم المادي والمعنوي.
- القيود المتعلقة بحقوق النشر.



ويُرجح العجمي (Al-Ajmi 2011) السبب في ندرة المدونات المتوازية التي تتضمن اللغة العربية إلى قلة البرمجيات اللازمة لتطوير مثل هذه المصادر، بالإضافة إلى افتقار المشاريع للدعم المؤسسي، والحكومي، والشكوك التي قد تساور بعض الباحثين حول جدوى استخدام مثل هذه المصادر.

وتتناول الأدبيات عدداً محدوداً من المشاريع التي تتضمن اللغة العربية (انظر الجدول ١)، وقد قمنا بتلخيص أكثرها تداولاً. ونبدأ بـ EAPCOUNT المدونة المتوازية لنصوص الأمم المتحدة، التي بدأت كأطروحة دكتوراه في جامعة (قرطاج) في تونس. تحتوي المدونة على ٢٦١ نصاً باللغة الإنجليزية (٧٥,٦٠٦ كلمة)، تشمل القرارات، والتقارير السنوية، والوثائق الصادرة عن منظمات الأمم المتحدة ومختلف مؤسساتها، تمت محاذاتها على مستوى الفقرة بترجمتها العربية. تحتوي المدونة في المجمل على ٥,٣٩٢,٤٩١ كلمة (انظر صالحي Salhi 2013). وفي مشروع مشابه تبنى مركز أبحاث الذكاء الاصطناعي في ألمانيا مشروع مدونة متوازية متعددة اللغات، من ضمنها اللغة العربية، تحت اسم MultiUN (انظر إيزلي وتشين Eisele, and Chen 2010). وتحتوي المدونة على ٣٠٠ مليون كلمة جُمعت من نصوص مترجمة لمنشورات الأمم المتحدة بمختلف أنواعها.

ومن المدونات المتوازية الشهيرة: المدونة المتوازية المفتوحة OPUS (انظر تيدمان Tiedemann 2012)، وهي مدونة مفتوحة المصدر، ومتعددة اللغات، تم فيها جمع النصوص المتوفرة على الشبكة العنكبوتية. كما تتضمن واجهة وأدوات حاسوبية مفتوحة المصدر، مما يجعلها مصدراً لغوياً فعالاً للباحثين والمهتمين بالدراسات اللغوية. وقد جمعت هذه المدونة بطريقة آلية تماماً، ولم تخضع للمراجعة البشرية، مما يعني وجود نسبة من الخطأ؛ سواء فيما يتعلق بجودة الترجمة، أو دقة المحاذاة. وتعد يورو ماتريكس EuroMatrix مدونة متوازية متعددة اللغات تم تمويلها من قبل الاتحاد الأوروبي، وتضم محاضر اجتماعات البرلمان الأوروبي مترجمة إلى العربية، والعديد من اللغات الأخرى. وتتضمن المدونة في مجملها ٥١ مليون كلمة، منها مليون ونصف المليون كلمة عربية تهدف إلى دعم البحث في مجال الترجمة الآلية.

المدونة	الجهة	المادة	الحجم	الغرض	المحتوى
المدونة المتوازية لنصوص الأمم المتحدة EAP- (COUNT	جامعة قرطاج	مكتوبة	٢٦١ نصاً مترجماً	دعم الأبحاث في مجالات اللغويات النظرية والتطبيقية	القرارات والتقارير السنوية، والوثائق الصادرة عن منظمات الأمم المتحدة، ومختلف مؤسساتها
المدونة المتوازية المفتوحة (OPUS)	مفتوحة المصدر	مكتوبة		دعم الأبحاث في مجالات اللغويات النظرية والتطبيقية، والترجمة الآلية	نصوص تم جمعها ومعالجتها آلياً من الشبكة العنكبوتية
مدونة الأمم المتحدة Mul- ((tiUN	أبحاث الذكاء الاصطناعي بألمانيا	مكتوبة	٣٠٠ مليون كلمة	لدعم البحث في مجال الترجمة الآلية	منشورات الأمم المتحدة
مدونة الاتحاد الأوروبي EuroMa- ((trix	الاتحاد الأوروبي	مكتوبة	٥١ مليون كلمة منها ١,٥ مليون كلمة باللغة العربية	لدعم البحث في مجال الترجمة الآلية	محاضر اجتماعات الاتحاد الأوروبي



المدونة	الجهة	المادة	الحجم	الغرض	المحتوى
مدونة جامعة الكويت	جامعة الكويت	مكتوبة	٣ ملايين كلمة	دعم الأبحاث في مجالات اللغويات، والدراسات المعجمية	نصوص مترجمة من سلسلة عالم المعرفة الصادرة من المجلس الوطني للثقافة والفنون والآداب في الكويت
مدونة ومنصة ترجمة الأفلام التعليمية (AMARA)	معهد قطر لبحوث الحوسبة	مسموعة	٢,٦ مليون كلمة بالعربية و٣,٩ مليون كلمة بالإنجليزية	لدعم البحث في مجال الترجمة الآلية	ترجمات لعدد من ملفات الفيديو التعليمية
المدونة العربية الإنجليزية	جامعة بانجور ببريطانيا	مكتوبة	٢٧,٨ مليون كلمة باللغة العربية و٣٠,٨ مليون	لاختبار نظم الترجمة الآلية	نصوص مترجمة من جريدة الحياة، ومن المدونة المتوازية المفتوحة OPUS
مدونة الأحاديث النبوية	جامعة ليدز ببريطانيا	مكتوبة	٣ مليون كلمة		أحاديث نبوية وترجماتها باللغات الإنجليزية والفرنسية والروسية

الجدول (١): ملخص لأشهر المدونات المتوازية والتي تتضمن اللغة العربية.



وقد تبَنّى اتحاد البيانات اللغوية Linguistic Data Consortium (LDC) في جامعة بنسيلفانيا Pennsylvania عدداً من المشاريع التي تتضمن اللغة العربية، وعن طريق عدة مراحل (انظر موقع الاتحاد www.ldc.upenn.edu). ومن أشهرها (GALE Phase 1)، حيث تم تطوير مدونة البرامج الإخبارية باللغتين (العربية والإنجليزية) بالمرحلة الأولى، وتحتوي نصوصاً من ١٧ ساعة بث إخباري، تم تجميعها من وكالات الأخبار العربية، ومواقع وبرامج إخبارية باللغة العربية، وترجمتها للغة الإنجليزية، ودُشّنت المدونة في عام ٢٠٠٧.

ويعاب على المدونات المتوازية التابعة لاتحاد البيانات اللغوية (LDC) ارتفاع أسعارها، وخاصة على الطلاب والباحثين، مما يحدّ من استخدامها على نطاق واسع، إلا أنها بالمقابل تمتاز بمعايير تصميمها العالية، ودقة وجودة الترجمة فيها. ويعرض الجدول (٢) قائمة بالمدونات المتوازية التي طورها اتحاد البيانات اللغوية، وعدد الكلمات في كل مدونة، بالإضافة إلى رسوم الاستخدام لغير أعضاء الاتحاد.

المدونة	عدد الكلمات	الرسوم بالدولار الأمريكي
ISI Arabic/English automatically extracted parallel text	٣١ مليون	٤٠٠٠ دولار
Uma Arabic/English parallel news pext	٢ مليون	٣٠٠٠ دولاء
Arabic/English parallel translation	٤٢,٠٠٠	٣٠٠٠ دولار
Multiple translation Arabic (part 1)	٢٣,٠٠٠	ألف دولار
Multiple translation Arabic (part 2)	١٥,٠٠٠	ألف دولار



المدونة	عدد الكلمات	الرسوم بالدولار الأمريكي
Arabic newswire English transaltion collection	٥٥١,٠٠٠	١٥٠٠ دولار
Arabic news translation text (part 1)	٤٤١,٠٠٠	٣٠٠٠ دولار
GALE phase 1 Arabic broadcast news parallel text (part 1)	٩٠,٠٠٠	١٥٠٠ دولار
IGALE phase 1 Arabic broadcast news parallel text (part 2)	٥٦,٠٠٠	١٥٠٠ دولار
UN bidirectional multi-lingual	مليون	٤٠٠٠ دولار

الجدول (٢): قائمة بالمدونات المتوازية التي طورها اتحاد البيانات اللغوية (مقتبس من القحطاني وتيهان Alkahtani and Teahan ٢٠١٥).

وفي مشروع آخر؛ تم تجميع مدونة متخصصة لنصوص تتناول تقنية المعلومات، تتضمن اللغة العربية، والإنجليزية، والسويدية (انظر إزفيني Izwaini 2003). تحتوي المدونة على (٧) ملايين كلمة باللغة الإنجليزية، ومليون كلمة باللغة العربية، و(٢,٧) كلمة باللغة السويدية. وقد تم تجميع النصوص من كتب، ودوريات علمية، ومواقع إلكترونية متخصصة بالحاسب، بالإضافة إلى أدلة الاستخدام المتوفرة لبعض أنظمة التشغيل الشهيرة، مثل (ويندوز ٩٨) وغيرها.

وفي مشروع آخر تبنته جامعة الكويت، تم تجميع مدونة متوازية تصل إلى ٣ ملايين



كلمة (انظر العجمي Al-Ajmi 2011). تحتوي المدونة على نصوص مترجمة من اللغة الإنجليزية للعربية، جُمعت من منشورات عالم المعرفة، وهي سلسلة كتب ثقافية تصدر شهرياً من المجلس الوطني للثقافة والفنون والآداب في دولة الكويت. وتهدف المدونة إلى خدمة الطلاب، والباحثين؛ حيث تم تصميم واجهة حاسوبية لاستعراض محتويات المدونة متاحة فقط لطلاب وأعضاء هيئة التدريس في جامعة الكويت، يمكنهم الدخول لها عن طريق حساباتهم الجامعية.

وفي مشروع بحثيٍّ حديث نسبياً أشرف عليه معهد قطر لبحوث الحوسبة، التابع لجامعة حمد بن خليفة، تم تطوير منصة لترجمة الفيديو التعليمي باسم AMARA، وتهدف هذه المنصة إلى دعم الترجمة الآلية للمحتوى التعليمي المرئي؛ حيث تضم مدونة متوازية تحتوي على ٢,٦ مليون كلمة باللغة العربية، و٣,٩ مليون كلمة باللغة الإنجليزية، تتضمن ترجمات لعدد من ملفات الفيديو التعليمية المتوفرة عبر منصات تعليمية شهيرة، مثل أكاديمية خان Khan Academy وتيد TED وغيرها (انظر عبد العالي وآخرين Abdelali et al. 2014).

ومن المشاريع الحديثة في مجال المدونات المتوازية والتي تتضمن اللغة العربية، قام الباحثان القحطاني وتيهان (Alkahtani & Teahan 2015) في جامعة بانجور في بريطانيا مؤخراً بتطوير مدونة متوازية تحتوي على نصوص عربية مترجمة للغة الإنجليزية، تحتوي المدونة على ٢٧,٨ مليون كلمة باللغة العربية، و٣٠,٨ مليون كلمة باللغة الإنجليزية، تغطي مجالات مختلفة من مصدرين أساسين: جريدة الحياة، والمدونة المتوازية المفتوحة OPUS. وتهدف المدونة إلى توفير بيانات لاختبار نظم الترجمة الآلية، ودعم الأبحاث في هذا المجال.

وفي مشروع حديث جداً، قامت جامعة (ليدز) بتصميم مدونة متوازية متعددة اللغات، تختص بنصوص الأحاديث النبوية الشريفة الواردة في صحيح البخاري باللغة العربية، وترجماتها باللغات: الإنجليزية، والفرنسية، والروسية، فيما يصل إلى مليوني كلمة (انظر حسان وأتويل Hassan and Atwell 2016).

وبعد استعراض مشاريع المدونات المتوازية التي تتضمن اللغة العربية في الجزء السابق؛ يتضح وجود حاجة للمزيد من الجهود في هذا المجال؛ حيث إن أغلب



المدونات التي استعرضناها إما غير متوفرة للطلاب والباحثين، أو تتطلب رسوماً عالية نسبياً. كما أن بعضها يُعدّ متخصصاً، ولا يغطي موضوعات مختلفة، أو أوعية نشر متعددة. ومن أوجه القصور كذلك في بعض هذه المدونات اعتماد بعضها على الترجمة الآلية، مما يعني تدني جودة الترجمة فيها، ومن هنا تبرز أهمية وجود مدونة متوازية للغتين (العربية والإنجليزية) تتناول موضوعات متعددة، وتشمل أوعية نشر مختلفة، يراعى فيها جودة الترجمة، ومستواها. كما ولابد من توفير واجهة حاسوبية سهلة الاستخدام، ومجانية تمكِّن الطلاب والباحثين من استعراض محتويات المدونة بسهولة، ويسر، بدون الحاجة لدفع أي رسوم.

نتناول في المحطة القادمة مشروع تطوير المدونة المتوازية AEPC التابعة لكلية اللغات والترجمة بجامعة الملك سعود ويشمل ذلك مراحل تصميم المدونة وخصائصها.

٥ .٦. مشروع AEPC

يهدف مشروع المدونة المتوازية الإنجليزية/ العربية AEPC إلى تصميم مدونة متوازية للنصوص الإنجليزية المترجمة للغة العربية. ولأن الهدف الأساسي للمدونة الاستفادة منها كمصدر لتعليم اللغة وتدريب المترجمين؛ فلابد من مراعاة أن تحقق الشروط التالية (انظر بايبر Biber 2003):

أن تكون ذات حجم ملائم: وتتضمن المرحلة الأولى لتطوير المدونة جمع (١٠) ملايين جملة باللغة الإنجليزية، تمت محاذاتها بترجمتها باللغة العربية.

الشمول (Representativeness)؛ ويعني ذلك أن تشتمل المدونة على نصوص تمثل استعمالات اللغة المختلفة، فلا تقتصر المدونة اللغوية على أسلوب معين أو منطقة جغرافية معينة مثلا.

التوازن (Balance)؛ ويعني ذلك ألا تقتصر المدونة على نصوص في تخصص معين، أو لكاتب معين، أو لزمن معين، إلخ.

الجودة (Quality)؛ ويقصد بذلك التأكد من جودة وسلامة الترجمة، ودقة المحاذاة.

الإتاحة (Availability) وهي أن تكون المدونة مجانية، ومتاحة للطلاب والباحثين عن طريق تطوير واجهة حاسوبية، وأدوات سهلة الاستخدام، تسمح باستعراض محتويات المدونة، والاستفادة منها.



وينقسم المشروع إلى مرحتين أساسيتين نوجزهما في المباحث القادمة.

٥. ٦. ١. المرحلة الأولى من مشروع AEPC

تتضمن المرحلة الأولى جمع النصوص، وتشذيبها، وتصميم الموقع التجريبي الذي يتضمن مكشافاً ثنائي اللغة لاستعراض محتويات المدونة. وتشمل المرحلة الثانية توسيم (ترميز) النصوص، وتطوير أدوات البحث في المدونة. وقد بدأت المرحلة الأولى للمشروع عام (٢٠١٥)، وتتضمن جمع وإعداد النصوص، وتصنيفها، وتجزئتها على مستوى الجملة، ومحاذاتها بترجماتها. كما تضمنت المرحلة الأولى تصميم واجهة حاسوبية ومكشافًا ثنائي اللغة للبحث في محتويات المدونة واستعراضها.

٥. ٦. ١. ١. إعداد النصوص

تمت الاستفادة من النصوص المترجمة في مختلف المجالات، والمتوفرة؛ إما إلكترونياً على الشبكة العنكبوتية، أو المطبوعة، وذلك بعد تحويلها إلى ملفات إلكترونية باستخدام برامج التعرف على الحروف OCR التي تدعم اللغة العربية، ومنها برنامج (ABBYY FineReader .12)، وبرنامج (Readiris .15).

وتتميز هذه البرامج بمستوى عال الدقة مقارنة بالبرامج والمواقع المشابهة، كما أن أسعارها لا تتجاوز (٩٩) دولارا. وقد تم جمع النصوص وتصنيفها وفق الفئات التالية:

اجتماعي Social/ سيرة ذاتية Biography/ أدبي Literary/ إداري Administrative/ طبي Medical/ قانوني Legal/ ديني Religious/ سياسي Political/ علمي Scientific.

وتحتوي المدونة على ملفات بصيغة (XML) وتتضمن البيانات الوصفية (Metadata) التالية باللغتين العربية والإنجليزية: (العنوان title)، و(المؤلف author)، و(الناشر publisher)، و(سنة النشر year of publication)، و(بلد النشر country)، و(جنس المؤلف author gender)، و(الوعاء medium)، و(المجال domain)، و(الموضوع topic).

٥. ٦. ١. ٢. التجزئة Segmentation

بعد جمع النصوص يتم تقسيمها إلى مقاطع قصيرة على مستوى الجملة، أو الفقرة. ويمكن استثمار بعض التطبيقات المتوفرة للقيام بهذه العملية آلياً مثل برامج ذاكرة



الترجمة كترادوس ستوديو Trados Stodio، أو برنامج ميم سورس Memsource، وغيرها، حيث تعمل هذه البرامج على تحميل أنواع مختلفة من الملفات النصية، وتعمل على تقسيم النص إلى مقاطع، وذلك بشكل آلي. إلا أن ذلك -بالطبع- لا يخلو من الأخطاء؛ حيث إن من الصعوبة أحياناً تحديد المقطع، أو بمعنى آخر تحديد ما يُعدّ جملة، فعلامات الترقيم مثل: النقطة، وعلامة التعجب، وعلامة الاستفهام تستخدم عادة للإشارة إلى نهاية الجملة، إلا أنها قد تخلق لهذه البرامج إشكالية؛ فالنقطة -على سبيل المثال- قد تعقب الاختصارات مثل Dr. أو Ms.، أو قد توجد بالأرقام العشرية، وعلى هذا فقد لا تعني بالضرورة نهاية جملة. (للمزيد عن برامج ذاكرات الترجمة؛ انظر العتيبي ٢٠١٦). وعليه؛ وبعد تجربة هذه البرامج، فقد تقرر أن تكون عملية التجزئة يدوية، وذلك لضمان دقة التجزئة وسلامتها من الأخطاء.

٥. ٦. ١. ٣. المحاذاة Alignment

تُعدّ المحاذاة -وتسمى في بعض المراجع بالمناظرة، أو المقابلة- مرحلة أساسية في إعداد نصوص المدونة المتوازية، وهي العملية التي يتم عن طريقها ربط الأجزاء الموجودة في النص الأصلي مع ترجماتها.

ويمكن للمحاذاة أن تتم على مستويات عدة: فمنها ما يكون على مستوى النص، ومنها ما يكون على مستوى الفقرة، أو الجملة، أو حتى على مستوى الكلمة، إلا أنه من الشائع أن تتم المحاذاة على مستوى الجملة، أو الفقرة.

ومثلما يمكن تنفيذ التجزئة آلياً؛ يمكن تنفيذ المحاذاة آلياً باستخدام بعض البرمجيات المتوفرة، مثل أداة وين ألاين WinAlign، أو يوبلغ UPlug، والتي تسمح برفع ملف النص الأصلي مع ملف الترجمة، بحيث يقوم البرنامج آلياً بالمرور على النص جملةً جملة، وربط الجملة الأولى من النص المترجم بالجملة الأولى من النص الأصلي، ثم الربط بين الجملة الثانية من النص المترجم والجملة الثانية من النص الأصلي، وهكذا دواليك بحيث ينتج في النهاية ملف بصيغة XML يحتوي على الجملة الأصلية، وبمحاذاتها ترجمتها.

وتتميز برامج المحاذاة الآلية بقدرتها على معالجة عدد كبير من الملفات بسرعة، كما يمكن أن تحقق نتائج معقولة حين يكون للنص الأصلي والنص المترجم بُنْيَة متشابهة -إن لم تكن متطابقة-، وهو مالا ينطبق على اللغتين العربية والإنجليزية. وتنطوي

عملية المحاذاة على العديد من التحديات، ومنها صعوبة ربط الجملة المستخدمة في النص الأصلي مع الجملة المترجمة؛ إذ من الممكن تقسيم جملة من لغة المصدر إلى جملتين في اللغة المترجم إليها، أو العكس، وعليه؛ فإن عدم وجود تطابق مباشر قد يخلق مشاكل لبرامج المحاذاة الآلية. ولهذا فقد تقرر بعد تجربة هذا البرامج اعتماد الطريقة اليدوية في عمل المحاذاة، وذلك لضمان دقتها وسلامتها من الأخطاء.

٥. ٦. ١. ٤. الواجهة والأدوات

من أهم أهداف هذا المشروع توفير مصدر تعليمي وبحثي للمهتمين بالترجمة، وتعليم اللغة تمكنهم من استعراض محتويات المدونة، والاستفادة منها مجاناً عن طريق



الشكل (١): واجهة موقع المدونة التجريبي والتي يمكن استعراض محتوياتها باللغتين العربية والإنجليزية.

٥. ٦. ٢. المرحلة الثانية من مشروع AEPC

تبدأ المرحلة الثانية بنهاية عام (٢٠١٦)، وتتضمن تحالفاً بحثياً مع كلية الحاسب والمعلومات بجامعة الملك سعود لتطوير أدوات المدونة، ويشمل ذلك أداة لعرض المتلازمات اللفظية (Colocations)، والتحليل الإحصائي، ومزيد من خيارات التصفية لعرض نتائج أكثر دقة. كما تشمل المرحلة الثانية توسيم النصوص بحسب نوع



الكلمة Part-of-speech tagging، ويقصد به إعطاء كل كلمة خصائصها وصفاتها وتصريفاتها النحوية حسب موقعها من الجملة، ومن المتوقع أن يساهم ذلك في تحسين نتائج البحث في المدونة. ومن أشهر برمجيات التوسيم حالياً: برنامج خوجة(Khoja 2001)، وبرنامج مداميرا (بيشا وآخرون Pasha et al 2014).

٥ . ٧. الخاتمة

قدمنا في هذا الفصل عرضاً موجزاً لتعريف القارئ العربي بالمدونات المتوازية؛ حيث تناولنا استخداماتها في تعليم اللغات، وفي تدريب المترجمين، ثم استعرضنا أشهر المدونات المتوازية التي تتناول اللغة العربية، ومن ثم قدمنا وصفاً لمشروع AEPC، وهو مشروع المدونة المتوازية (الإنجليزية-العربية) التابعة لكلية اللغات والترجمة في جامعة الملك سعود، والذي يهدف إلى توفير مصدر لغوي، وتعليمي، يسهم في دعم البحث في مجالات اللسانيات، والترجمة، وتعليم اللغات.

قائمة المراجع

أولاً: المراجع العربية

- البواب، مروان. (٢٠٠٩)، أثر التقانات الحديثة في تجديد المعجم العربي. دراسة مقدمة إلى المؤتمر السنوي السابع لمجمع اللغة العربية بدمشق.
- توتاوي، فاطمة الزهراء. (٢٠١٠)، استثمار المدونة في تدريب المترجم. رسالة دكتوراه- جامعة وهران.
- الثبيتي، عبدالمحسن عبيد. (٢٠٠٧)، استخدام ذخائر النصوص لاستخلاص المصطلحات المتخصصة» في: الندوة الدولية الأولى عن الحاسب واللغة العربية: الأوراق البحثية، ص ٣١-٣٨.
- جمعة، عمرو. (٢٠١٥)، صناعة المناهج وتطويرها باستخدام التقنيات الحديثة: نحو منهج معاصر للغة العربية. مؤتمر كيوساينس (٢٠١٥)، ٤٠.
- حماده، سلوى. (٢٠٠٧)، نحو منهج مقترح لصناعة المدونات، المؤتمر السابع لجمعية هندسة اللغة, ٤-٥ سبتمبر، ٢٠٠٧، القاهرة.

- صالح، محمود إسماعيل. (٢٠٠٩)، الحاسوب في خدمة المترجمين، المؤتمر الترجمة والتعريب في المملكة العربية السعودية، الرياض.
- صالح، محمود إسماعيل. (٢٠١٢)، الحاسوب في البحث اللغوي: لسانيات المدونات اللغوية أنموذجا. الرياض: جامعة الأميرة نورة.
- العتيبي، هند مطلق. (٢٠١٦)، نظم ذاكرات الترجمة، مجلة اللسان العربي، العدد ٧٦. مكتب تنسيق التعريب، الرباط.
- العصيمي، صالح فهد. (٢٠١٣)، لسانيات المتون وعلوم اللغة، في: مجلة كلية الآداب والعلوم الإنسانية (فاس، المغرب) العدد ١٩، السنة الخامسة والثلاثون، ص ٣٧-٦٧.
- الفيفي، عبدالله يحي. (٢٠١٢)، المدونات اللغوية لمتعلمي اللغة العربية: نظام لتصنيف وترميز الأخطاء اللغوية. في: السجل العلمي للمؤتمر الدولي لعلوم وهندسة الحاسوب في اللغة العربية في دورته الثامنة (٢٦-٢٨ ديسمبر، ٢٠١٢) جامعة القاهرة.
- مكانري، توني وهاردي، أندرو. (٢٠١٦). لغويات المدونة الحاسوبية: المنهج والنظرية والتطبيق. دار جامعة الملك سعود للنشر، الرياض.
- المجيول، سلطان. لغويات المدوّنة العربية الحاسوبية. جريدة الجزيرة، ٢٩ أغسطس ٢٠١٥. مستردة بتاريخ ١٢ أكتوبر، ٢٠١٦.
http://www.al-jazirah.com/2015/20150829/cu15.htm.

ثانيًا: المراجع الأجنبية

- Aijmer, K. (Ed.). (2009). Corpora and language teaching (Vol. 33). Amsterdam and Philadelphia: John Benjamins Publishing.
- Al-Ajmi, H. (2003). Compiling an English-Arabic parallel text corpus. In M. Murata, S.Yamada & Y. Tono (Eds), Proceedings of Asian Association for Lexicography (pp.51–54).
- Al-Ajmi, H. (2004). A new English–Arabic parallel text corpus for lexicographic applications. Lexikos,14(1).
- Alkahtani, S., and Teahan, W. J. (2015, December). A New Parallel Corpus of Arabic/English. In Proceedings of the Eighth Saudi Students Conference in the UK (p. 279). World Scientific.



Almujaiwel. S. (2012). Contrastive Lexicology and Comparable English-Arabic Corpora-Based Analysis of Vague and Mistranslated Arabic Equivalence. Ph.D. thesis. Exeter: The University of Exeter.

Aston, G. (1999). Corpus use and learning to translate. Textus, 12(2), pp. 289– 314.

Baker, M. (1993). Corpus linguistics and translation studies: Implications and applications. In M. Baker, G. Francis, and E. Tognini-Bonelli (Eds.), Text and technology: In honour of John Sinclair. Amsterdam and Philadelphia: John Benjamins, pp. 233-250.

Baker, M. (1995). Corpora in translation studies: An overview and some suggestions for future research. Target, 7(2), 223-243.

Baker, M. (1998). Routledge encyclopedia of translation studies. London: Routledge: 74-76, 244-248.

Baker, M. (1999). The role of corpora in investigating the linguistic behaviour of professional translators. International Journal of Corpus Linguistics, 4(2), 281-298.

Barlow, M. (1996). Parallel texts in language teaching. In S. Botley, J. Glass, T. McEnery, & A. Wilson (Eds.), Proceedings of teaching and language corpora 1996 (pp. 45-56). Lancaster, UK: UCREL Technical Papers Volume 9.

Bernardini, S. (2004). Corpus-aided language pedagogy for translation education. In K. Malmkjær (Ed.), Translation in undergraduate degree programmes (pp. 97-111). Amsterdam and Philadelphia: John Benjamins Publishing.

Bernardini, S. and Castagnoli, S. (2008). Corpora for translator education and translation practice. In E. Yuste-Rodrigo (Ed.), Topics in language resources for translation and localization (pp. 39-55). Amsterdam and Philadelphia: Benjamins Publishing.

Biber, D. (1993). Representativeness in corpus design. Literary and linguistic computing, 8(4), 243-257.

Botley, S., Glass, J., McEnery, A. and Wilson, A. (1996). Proceedings



of TALC 96, Lancaster, UK: UCREL Technical Papers Volume 9.

- Botley, S. P., McEnery, A. M., and Wilson, A. (Eds.). (2000). Multilingual corpora in teaching and research. Amsterdam-Atlanta, G.A.: Rodopi

- Bowker, L., and Pearson, J. (2002). Working with specialized language: a practical guide to using corpora. London: Routledge.

- Brown, P. F., Pietra, V. J. D., Pietra, S. A. D., and Mercer, R. L. (1993). The mathematics of statistical machine translation: Parameter estimation. Computational linguistics, 19(2), 263-311.

- Brown, P. F., Lai, J. C., and Mercer, R. L. (1991, June). Aligning sentences in parallel corpora. In Proceedings of the 29th annual meeting on Association for Computational Linguistics (pp. 169-176). Association for Computational Linguistics.

- Chujo, K., Anthony, L., and Oghigian, K. (2009). DDL for the EFL classroom: Effective uses of a Japanese-English parallel corpus and the development of a learner-friendly, online parallel concordancer. In Proceedings of the corpus linguistics conference, CL2009. Liverpool University of Liverpool.

- Cohn, T. and Lapata, M., (2007). Machine translation by triangulation: Making effective use of multi-parallel corpora. In Proceedings of 45th Annual Meeting of the Association for Computational Linguistics, pages 828-735.

- Davies, G., Otto, S. and Rüschoff, B. (2013). Historical perspectives on CALL. In: M. Thomas, H. Reinders & M. Warschauer (a c. di.), Contemporary computer-assisted language learning (pp. 19-38). Londra: Bloomsbury.

- Eisele, A., and Chen, Y. (2010). MultiUN: A Multilingual Corpus from United Nation Documents. In Proceedings of the 7th International Conference on Language Resources and Evaluation (LREC). European Language Resources Association (ELRA), pp. 2868– 2872.

- Flowerdew, L. (2015). Data-driven learning and language learning



theories: Whither the twain shall meet. In A. Leńko-Szymańska and A. Boulton (Eds.), Multiple affordances of language corpora for data-driven learning (pp. 15–36). Amsterdam: John Benjamins.

- Frankenberg-Garcia, A. (2015). Training translators to use corpora hands-on: challenges and reactions by a group of thirteen students at a UK university. Corpora, 10(3), 351-380.

- Frankenberg-Garcia, A., and Santos, D. (2003). Introducing Compara: The Portuguese -English parallel corpus. Corpora in translator education, 71-87.

- Frankenberg-Garcia, A., Flowerdew, L., and Aston, G. (Eds.). (2011). New trends in corpora and language learning. London: Continuum.

- Godwin-Jones, B. (2001). Emerging technologies: tools and trends in corpora use for teaching and learning. Language, Learning & Technology, 5(3), 7-7.

- Gouadec, D. (2002). Training translators: certainties, uncertainties, dilemmas. Training the Language Services Provider for the New Millennium, Oporto: Universidade do Porto, 31-41.

- Granger, S., and Lefer, M. A. (2016). From general to learners' bilingual dictionaries: Towards a more effective fulfillment of advanced learners' phraseological needs. International Journal of Lexicography, ecw022.

- Guzman, F., Sajjad, H., Abdelali, A., and Vogel, S. (2013). The AMARA corpus: Building resources for translating the web's educational content. In Proceedings of the International Workshop on Spoken Language Translation, IWSLT '13, Heidelberg, Germany.

- Haiyan, L. and Zhongxin, D. (2014). Effectiveness of self-built Chinese-English Corpus on Assisting Translation Teaching. International Journal of Humanities and Social Science, 4(7), 96-98.

- Harris, B. (1988). Bi-text, a new concept in translation theory. Language Monthly, 54, 8-10.



Hassan, S., and Atwell, E. S. (2016). Design and Implementing of Multilingual Hadith Corpus. International Journal of Recent Research in Social Sciences and Humanities, 3(2), 100-104.

Johansson, S. (1999). Corpora and contrastive studies. Multiple Languages–Multiple Perspectives. AFinLA Yearbook, (57), 116-125.

Johansson, S. (2007). Seeing Through Multilingual Corpora: On the Use of Corpora in Contrastive Studies. Amsterdam and Philadelphia: John Benjamins.

John, E. S. (2001). A case for using a parallel corpus and concordancer for beginners of a foreign language. Language learning & technology, 5(3), 185-203.

Johns, T. (1986). Micro-concord: A language learner›s research tool. System,14(2), 151-162.

Johns, T. (1993). Data-driven learning: An update. TELL & CALL, 3, 23-32.

Johns, T. 2002. Data-driven learning: The perpetual challenge. In Kettermann, B. and Marko, G. (eds.) Teaching and learning by doing corpus linguistics. Amsterdam: Rodopi, 107-117.

Khoja, S. (2001, June). APT: Arabic part-of-speech tagger. In Proceedings of the Student Workshop at NAACL (pp. 20-25).

Kilgarriff, A. (2013). Terminology finding, parallel corpora and bilingual word sketches in the Sketch Engine. In Proceedings OF ASLIB 35th Translating and the Computer Conference, London.

Krishnamurthy, R. (2000). Size Matters: Creating Dictionaries from the World's Largest Corpus. In Proceedings of KOTESOL 2000 Casting the Net: Diversity in Language Learning. Taegu, Korea.

Lee, J. H., Lee, H., and Sert, C. (2015). A Corpus Approach for Autonomous Teachers and Learners: Implementing an On-Line Concordancer on Teachers› Laptops. Language Learning & Technology, 19(2), 1-15.

Leech, G.N.)2010(. Corpus linguistics. In K. Malmkjaer (ed.), The



Routledge Linguistics Encyclopedia, Third edition. London: Routledge, 103-113.

- Linguistic Data Consortium. 2013. LDC Catalog. [Online]. Available at: http://catalog.ldc.upenn.edu

- Marco, J., and van Lawick, H. (2015). Enhancing translator trainees' awareness of source text interference through use of comparable corpora. Multiple Affordances of Language Corpora for Data-driven Learning, 69, 225.

- Olohan, M.)2004(. Introducing corpora in translation studies. London: Routledge.

- Pasha, A., Al-Badrashiny, M., Kholy, A. E., Eskander, R., Diab, M., Habash, N., Pooleery, M., Rambow, O., and Roth, R. (2014). MADAMIRA: A Fast, Comprehensive Tool for Morphological Analysis and Disambiguation of Arabic. In Proceedings of the 9th International Conference on Language Resources and Evaluation, Reykjavik, Iceland

- Pearson J. (2003). Using parallel texts in the Translator Training Environment. In Zanettin F., Bernardini S. and Stewart D. editors, Corpora in Translator Education. Manchester: St Jerome Publishing

- Perez, M. M., Paulussen, H., Macken, L., and Desmet, P. (2014). From input to output: the potential of parallel corpora for CALL. Language Resources and Evaluation, 48(1), 165-189.

- Rodríguez-Inés, P. (2014). Using corpora for awareness-raising purposes in translation, especially into a foreign language (Spanish–English). Perspectives, 22(2), 222-241.

- Römer, U. (2006). Pedagogical applications of corpora: Some reflections on the current scope and a wish list for future developments. Zeitschrift für Anglistik und Amerikanistik, 54(2), 121-134.

- Salhi, H. (2013). Investigating the Complementary Polysemy and the Arabic Translations of the Noun Destruction in EAPCOUNT. Meta: Journal des aducteursMeta:/Translators' Journal, 58(1), 227-246.



- SANGAWA, K. H. (2014). The learner as lexicographer: using monolingual and bilingual corpora to deepen vocabulary knowledge. Acta Linguistica Asiatica, 4(2), 53-65.

- Sharoff, S., Rapp, R., Zweigenbaum, P., and Fung, P. (Eds.). (2013). Building and Using Comparable Corpora. Springer.

- Sinclair, J. (1991). Corpus, concordance, collocation. Oxford: Oxford University Press.

- Sinclair, J. (1995). Corpus typology: A framework for classification. Stockholm studies in English, 85, 17-33.

- Sinclair, J.M. ed.)2004(. How to use corpora in language teaching. Amsterdam: John Benjamins Publishing.

- Stevens, V. (1991). Classroom concordancing: Vocabulary materials derived from relevant, authentic text. English for Special Purposes Journal, 10, 35-46.

- Teubert, W. (2002). The Role of Parallel Corpora in Translation and Multilingual Lexicography, in Altenberg, B. and S. Granger (eds.) Lexis in Contrast, pp. 189–214. Amsterdam: John Benjamins.

- Tiedemann, J. (2012). Parallel Data, Tools and Interfaces in OPUS. In Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC'12).

- Tiedemann, J. (2003). Recycling Translations Extraction of Lexical Data from Parallel Corpora and their Applications in Natural Language Processing. PhD Thesis. Uppsala University.

- Verplaetse, H., and Heylen, K. (2015). Parallel Corpora for Medical Translation Training: An Analysis of Impact On Student Performance. In proceedings of the 4th International Conference on Corpus Use and Learning to Translate (CULT). Alicante, Spain.

- Wichmann, A., and Fligelstone, S. (2014). Teaching and language corpora. London: Routledge.

- Wong, T. S., and Lee, J. S. (2016). Corpus-based learning of Cantonese for Mandarin speakers. ReCALL, 28(02), 187-206.


هذه الطبعة
إهداء من المركز
ولا يسمح بنشرها ورقياً
أو تداولها تجارياً

الفصل السادس

د. عقيل بن حامد الشمري
د. محمود بن عبدالله المحمود

المعالجة الآلية لوعاء الأخبار: تحليل الخطاب النقدي المعتمد على المدونة الحاسوبية

الملخص

اعتمد هذا الفصل في تحليل نصوص الأخبار على مدونة لغوية من قناتي العربية والجزيرة حول أخبار الحرب على غزة في النصف الثاني من العام ٢٠١٤م. وبلغ عدد الكلمات الكلي للمدونة المدروسة ٥٢٨٦٤١ كلمة، موزعة على ١٥٢٥ نصا إخباريا، تمثل نصوص قناة العربية منها: ٣١٪، بينما تمثل نصوص قناة الجزيرة ٦٩٪. وأُجري التحليل الكمي لمدونة الدراسة عن طريق استخراج قوائم التكرار للكلمات، واستخراج الكلمات المميزة لنصوص كل قناة على حدة، واستعراض الكشاف السياقي، وتحليل الكلمات المصاحبة للكلمات الأكثر تكرارا وتميّزا من حيث التوارد. وصُنّفت الكلمات عن طريق استقراء بيانات الدراسة إلى حقول دلالية تعكس طبيعتها. وأظهرت الدراسة وجود تشابه بين القناتين في بعض الجوانب، وهذه الجوانب هي: التركيز على الكلمات التي تقدم الخبر وتحكي الحدث ومن ثمّ تصف الحراك السياسي ذي العلاقة. ومن الجانب الآخر أظهرت الدراسة وجود اختلاف واضح بين قناتي الدراسة في جوانب أخرى؛ من أبرزها: الاختلاف في الإطار العام لتقديم الحدث، ونوعية الكلمات المستخدمة لتوصيف بعض جوانبه. فالعربية تميل إلى التركيز على الرؤية الرسمية الدولية، بينما يظهر في الجزيرة اهتمامٌ بالوصف والسرد التفصيلي للوقائع وما يحيط بها من ملابسات.



وظهر أيضا في العربية محاولة انتقاء الكلمات العمومية، بينما تشير المدونة إلى استخدام الجزيرة لمفردات مركزية تتضمن حمولات ثقافية وتاريخية، وتعبر عن موقف توصيفي تجاه الأحداث.

الكلمات المفتاحية: المعالجة الآلية للأخبار، التحليل النقدي للخطاب، تحليل الخطاب الإعلامي، المدونات الحاسوبية، لسانيات المدونات.

٦. ١. المقدمة[١]

يَعتَبِرُ التحليل النقدي للخطاب حيادية اللغة أمرًا لا وجود له. فاللغة مندمجة في ارتباطات اجتماعية معينة، وهي تعكس تلك الارتباطات بدرجات مختلفة سواء وعى ذلك مستخدم اللغة أم لم يعه. وتنطبق الرؤية السابقة على لغة الصناعة الإعلامية بوجه عام. وربما تتضح بصورة أكبر من خلال الأخبار وصياغتها اللغوية التي تمثل حجر الزاوية في بناء رؤية –واعية أم غير واعية– تنعكس في التعبير عن الوقائع وفي نقل الأحداث. ويحاول التحليل النقدي للخطاب الكشف عن الروابط الظاهرة والخفية بين اللغة من ناحية، والبنى السياسية والاجتماعية التي تنتمي إليها وتمثلها من ناحية أخرى، بدلا من الاكتفاء بوصف أنماط الاستعمال اللغوي وكأنها اختيارات محايدة أو مواضعات اعتباطية بحتة.

وفي الدراسة الحالية يسعى الباحثان إلى التحليل النقدي للخطاب من خلال استثمار المدونة اللغوية المتعلقة بأخبار الحرب على غزة في النصف الثاني من عام ٢٠١٤م في قناتي العربية والجزيرة. وتبتدئ الورقة البحثية بإطار نظري مكثف لمناقشة تحليل الخطاب بوجه عام، والتحليل النقدي للخطاب بوجه أخص مرورا بالخطاب السياسي في الإعلام، واستخدام المدونات اللغوية في تحليل الخطاب. وبعد استعراض الدراسات

١– يتقدم الباحثان بجزيل الشكر والامتنان إلى د. عبد المحسن بن عبيد الثبيتي (أستاذ البحث المشارك بالمركز الوطني لتقنية الحاسب والرياضيات التطبيقية بمدينة الملك عبدالعزيز للعلوم والتقنية، والمشرف على مشروع المدونة اللغوية العربية لمدينة الملك عبد العزيز للعلوم والتقنية، السعودية) الذي زوّدنا بأدوات التحليل، وبيان طريقة استخدامها، وإبداء الملاحظات على آلية جمع البيانات والعمليات المتعلقة بها. كما يتقدم الباحثان بجزيل الامتنان إلى د. سلطان المجيول (محرر الكتاب) على دعوته للمشاركة في الكتاب، وتفضله بالقراءة الفاحصة للبحث، ومساهمته في تحريره بشكل فاعل. وينوه الباحثان إلى أن أصل هذا الفصل سبق نشره في مجلة الحكمة للدراسات الأدبية واللغوية، العدد الثالث والثلاثون، ٢٠١٥م، غير أن النسخة الحالية قد عُدلت وفقا لمعيارين في حقل لسانيات المدونات. المعيار الأول: التناتجية replicability حيث أعاد الباحثان التحليل التجريبي واختبرا النتائج بمزيد من الوضوح. المعيار الثاني: معاجلة مدونتي الدراسة مرة أخرى من أجل مبدأ التعليلية accountability (حول هذه المعايير؛ انظر: مكانري وهاردي ٢٠١٢: ٢٦–٣٠).



السابقة التي تعالج المضامين السياسة في وسائل الإعلام من زاوية تحليل الخطاب؛ تنتقل الدراسة لاستعراض منهجية البحث من خلال وصف بيانات الدراسة ومحدداتها الزمانية والموضوعية، والأدوات المستخدمة في تحليل المدونة. وفي المبحث الذي يليه يستعرض الباحثان أبرز نتائج الدراسة من خلال استعراض الكلمات الأكثر تكرارا والكلمات المميزة والمحاور الدلالية التي تنتمي لها وأبرز مصاحباتها اللفظية. وفي الشق الأخير يناقش الباحثان نتائج الدراسة من خلال أوجه الشبه والاختلاف بين القناتين عينة الدراسة مع ربط ذلك بأدبيات الدراسة.

٦. ٢. تحليل الخطاب

يصعب جدا تحديد المقصود بـ (تحليل الخطاب) بشكل مختصر ودقيق، فالمجال الذي يعبر عنه هذا المصطلح مجال واسع ومتشعب، وليس له نظرية موحدة ومستقرة. وتعود هذه الإشكالية في تعريف المجال بدرجة كبيرة إلى ظروف نشأته وتشكله حيث نشأ وتشكل عبر تضافر عدد من الروافد والتيارات المختلفة والمتباينة في اتجاهاتها وأولوياتها البحثية والمعرفية. ونشأ عن ذلك الكثير من الاضطراب والاختلاف في تحديد المقصود بـ (الخطاب) من ناحية، وتعريف المجال المعني بتحليل الخطاب من ناحية أخرى (للمزيد، انظر: Gee 2014، و Mills 1997)، بل يوجد من يشكك أصلا في المنزلة العلمية لتحليل الخطاب من الناحية الإبستمولوجية (انظر على سبيل المثال: ريبول وموشلير ٢٠١٤). وبطبيعة الحال، فإننا لن نتوقف طويلا عند التفاصيل المتعلقة بهذا الموضوع لأنها تقع خارج نطاق المقصود هنا، ولكننا نعتقد أنه من المهم أن نشير بإيجاز إلى أهم الاتجاهات المعرفية العامة في هذا المجال من أجل تحديد منطلق الدراسة الحالية وموضعها.

أما المنطلق الأساسي لتحليل الخطاب فيمكن إرجاعه إلى رافدين أساسيين: أحدهما فكري/ فلسفي، والآخر لساني. فتحليل الخطاب من الناحية الفلسفية مرتبط بالتحول الإبستمولوجي الذي شهده الفضاء الفكري العام منذ منتصف القرن العشرين في النظر إلى المعرفة، وطبيعتها، ودور اللغة في تشكيلها وتوجيهها (Chouliaraki 2008 و Jaworski & Coupland 2006). فخلافا للنظرة التقليدية التي تربط اللغة بالمنطق وتفترض أن اللغة تقوم بدور تمثيلي محايد نوعا ما ومطابق للواقع الخارجي عبر النقل المجرد للمعرفة الموضوعية، أصبح يُنْظر إلى اللغة على أنها وسيط له دور أساسي في



تشكيل المعرفة على نحو معين، حيث يتخلق المعنى داخل السياق التداولي بشكل اعتباطي وبمعزل عن الارتباط بأي بنية منطقية مفترضة مسبقا في الواقع الخارجي (لا يمكن التفصيل هنا، وللمزيد حول تحولات الفلسفة اللغوية في هذا الصدد يمكن الرجوع إلى: Burge 1992). وبطبيعة الحال، فإن هذا التوجه يضم بداخله العديد من التيارات المختلفة والمتباينة، ولكنها تشترك في التركيز على (الخطاب) باعتباره المفهوم المركزي الذي يدل على طريقة معينة في رؤية الواقع والتعبير عنه[١] (انظر Chouliaraki 2008؛ Jørgensen & Phillips 2002 و Mills 1997).

لقد أدى هذا التوجه الفكري إلى اهتمام العلوم الاجتماعية بدور اللغة باعتبارها مظهرًا للخطاب الأساسي الذي يجسّده ويعبر عنه ويضعه في موضع التنفيذ في الأشكال المختلفة للتعاملات الاجتماعية (Howarth 2000). ومع أهمية المنحى الفلسفي والفكري العام في تأسيس دراسات الخطاب، إلا أن تأثيره الأكبر كان محصورا في الدراسات الاجتماعية والثقافية خارج نطاق اللسانيات، لأن هذا التوجه في مجمله ذو صبغة عمومية وتجريدية لا تهتم كثيرا بالتحليل اللساني الدقيق والمفصل وما يناسبه من أدوات وآليات منهجية (فاركلوف ٢٠٠٩). ولكن اللسانيات، وفي نفس الفترة تقريبا، بدأت تشهد تحولات مماثلة ترتكز على العناية باستعمال اللغة فعلا في مقامات التواصل المختلفة، وما يتصل بذلك من وضعيات وارتباطات معرفية واجتماعية وسياسية ونحوها[٢] (Jaworski & Coupland 2006). وقد تمثل هذا التوجه بظهور عدد من

١- لا يتسع المقام في هذا الاستعراض الموجز جدا للوقوف عند كثير من التفاصيل، ولكن النقد الماركسي لمفهوم (الأيديولوجيا) وارتباطه الوثيق باللغة قد يكون أحد المنطلقات الأساسية في تشكيل المنعطف اللغوي ومن أكثر التيارات تأثيرا في التوجه نحو «الخطاب» ودراسته. فبحسب النقد الماركسي فإن الأيديولوجيا تتولد أساسا عبر ممارسة الكلام وإنتاج العلامات في معترك الحياة الاجتماعية. ولذا فإن فعل الكلام المتعين والمندمج في الحراك الاجتماعي ينبغي أن يكون هو موطن الدراسة اللغوية. وقد أدى هذا التصور منذ وقت مبكر إلى معارضة النظرة التجريدية التي تقوم على الفصل بين قطبي اللغة والكلام في التأسيس المعرفي لعلم اللسانيات كما بدأه دي سوسير (انظر: Vološinov ١٩٧٣). وقد يكون الناقد الفرنسي الشهير ميشيل فوكو من أبرز الذين تأثروا بهذا الاتجاه، وكان لهم دور كبير جدا في انتشار الاهتمام بالخطاب في مجال الدراسات الإنسانية والاجتماعية بشكل عام (انظر مثلا: Jørgensen & Phillips ٢٠٠٢). ولقد تسرب تأثير هذه التصورات تدريجيا وبأشكال مختلفة إلى الدراسات اللغوية واللسانية عبر التوجه الذي يدرس استعمال اللغة من منظور ما عرف لاحقا باسم «التحليل النقدي للخطاب» (انظر مثلا: Van Dijk ١٩٩٣). وسنشير إلى ذلك لاحقا.

٢- لابد من التوضيح هنا بأن اللسانيات في منطلقها الأساسي لم تكن معينة بالقضايا المتعلقة باستعمال اللغة، بل كانت تركز على دراسة البنية المجردة للنظام اللغوي وخصائصها الشكلية في مستوى الجملة تحديدا. فالفصل بين (اللسان) و(الكلام) أو بين اللغة باعتبارها نظاما شكليا مجردا من ناحية، واستعمال اللغة فعلا من قبل المتحدثين من ناحية أخرى،



الاتجاهات البحثية المتنوعة مثل: (نظرية أفعال الكلام)، و(التداولية)، و(أثنوغرافية التواصل)، و(تحليل التحادث)، بالإضافة إلى (تحليل التنوع اللغوي) (Beaugrande 1997 و Schiffrin 1994).

إن هذه الاتجاهات بمجموعها تشكل الرافد اللساني لمباحث تحليل الخطاب. فرغم ما بينها من اختلاف وتنوع، إلا أنها تشترك جميعا في التركيز على (استعمال اللغة)، وافتراض وجود تنظيم يحكم ذلك يتجاوز البنية الأساسية والشكلية للجملة، مع البحث عن الوحدات التحليلية المناسبة لهذا الغرض. وتبعا لذلك برز مفهوم (الخطاب) باعتباره أحد المفاهيم الأساسية لتحليل اللغة في موضع الاستعمال. وكما توضح دوبرا شيفرن (Schiffrin 1994)، فإن هذه الاتجاهات رغم اشتراكها في العناية بمفهوم الخطاب، إلا أنها تختلف في تعريف هذا المفهوم تبعا لما بينها من اختلاف في الأولويات والاهتمامات النظرية والبحثية. وكما تبين شيفرن أيضا، فإن تعريفات الخطاب في المباحث اللسانية رغم اختلافها وتنوعها ترجع إلى ثلاثة معان رئيسية لا تخرج عنها، ويمكن تلخيصها كالتالي:

أولاً: الخطاب بمعنى متوالية من الأقوال والجمل المتعاقبة والمترابطة فيما بينها بطريقة معينة لتشكيل وحدة لسانية قائمة بذاتها. والخطاب بهذا المعنى ذو مدلول مرادف تقريبا لمفهوم «النص»، ويقابل مفهوم «الجملة».

ثانيًا: الخطاب بمعنى استعمال اللغة فعلا فهما أو إنتاجا في مقام محدد مرتبط بنوعية معينة من الظروف والملابسات والاعتبارات التعاملية التي تشكل الحدث التواصلي وتوجه مساره بطريقة معينة. والخطاب بهذا المعنى ذو مدلول مرادف تقريبا لمفهوم «الكلام» الذي يقابل مفهوم «اللسان» تبعا للتميز التقليدي في اللسانيات بين هذين المفهومين.

ثالثًا: الخطاب بمعنى نمط خاص من الممارسات اللغوية التي تتضمن استعمال اللغة بطريقة معينة لأداء وظائف محددة تبعا لشروط واعتبارات اجتماعية متداخلة تحكم

بالإضافة إلى اعتبار الجملة وحدة التحليل الأساسية من الركائز الإبستمولوجية الراسخة في اللسانيات التقليدية تبعا لـ «دي سوسير» والاتجاهات التي ورثت هذا المنحى مثل البنيوية والتوليدية (انظر المقدمة الضافية التي كتبها د. عزالدين المجدوب في نشأة اللسانيات والتحولات المعرفية التي شهدها هذا المجال في مراحل تطوره المختلفة: مجدوب، ٢٠١٢). ولاتزال دراسة المسائل المتعلقة باستعمال اللغة خارج نطاق اللسانيات بمفهومها الدقيق والمضيق.



بمجموعها أنماط الاستعمال اللغوي وتجعله يأخذ صبغة معينة لها سماتها وخصائصها المميزة. والخطاب بهذا المعنى يستخدم بمدلول يشمل مفهوم «النص» بالإضافة إلى الخلفيات والارتباطات المعرفية والاجتماعية التي تمثل قواعد صياغة النص وتشكيله بطريقة معينة.

وواضح أنه رغم ما بين هذه المعاني من تفاوت طفيف نوعا ما، إلا أن بينها الكثير من التداخل. فكلها تتفق على أن الخطاب هو الأقوال التي تمثل استخدام اللغة فعلا في سياقاتها الاجتماعية، مع مراعاة ما تفرضه تلك السياقات من أشكال واستخدامات خاصة. والاختلاف بين هذه المعاني المتنوعة يعود بالدرجة الأولى إلى موطن التركيز الأساسي. فالمعنى الأول يركز على الخطاب باعتباره ناتجا منجزا بسمات معينة، وأما الثاني فيركز على الخطاب باعتباره عملية إنجاز في ظل ظروف معينة، وأما الثالث فيركز على الخطاب باعتباره ممارسة تشمل مجموع الأمرين وما بينهما من تفاعل أثناء أداء الوظائف اللغوية بشروط معينة تشمل على سبيل المثال: الانتماء إلى صنف من المتكلمين، أو إلى توجه أيديولوجي، أو حقل خطابي معين (وانظر أيضا: شارودو ومنغنو ٢٠٠٨: ١٨٠).

وتبعا لهذه التنوعات في تحديد مفهوم «الخطاب» هناك تنوعات مماثلة في تحديد المقصود بـ «تحليل الخطاب» في الدراسات اللسانية. فتتوزع الاتجاهات في هذا الصدد بين اتجاهات ذات نزعة نصية، وأخرى ذات توجه اجتماعي (Paltridge 2012؛ وانظر أيضا: فاركلوف، ٢٠٠٩). أما الاتجاهات ذات النزعة النصية فتتداخل مع ما يسمى بـ «لسانيات النص» (خطابي ٢٠١٢)، وتركز بالدرجة الأولى على النصوص وتنظيمها إما بتحليل خواصها الشكلية أو بدراسة حركة الوصل الرابطة بين تنظيم نصي معين ومقام اجتماعي معين. وأما الاتجاهات ذات المنزع الاجتماعي فبعضها يركز بالدرجة الأولى على العمليات التفاعلية المباشرة التي تتضمنها نشاطات التخاطب واستعمال اللغة فهما أو إنتاجا في مقامات معينة (انظر مثلا: براون ويول ١٩٩٧). وبعضها ذو توجه أشمل لأنه يضع عمليات استعمال اللغة في سياقاتها الاجتماعية العامة. ويحاول تحليل الخطاب باعتباره نشاطا اجتماعيا مرتبطا ارتباطا عضويا وثيقا بالتنظيمات والهياكل والمؤسسات الاجتماعية التي يشكّلها ويتشكل بها في نفس الوقت (Gee 2014 و van Dijk 1997).

وهذا الاتجاه الأخير يتجاوز الفصل بين التحليل اللساني والتحليل الاجتماعي ويدمج



بينهما لأنه يعتبر تحليل النصوص شكلا من أشكال التحليل الاجتماعي باعتبار الخطاب هو الأساس في تشكيل الهوية والمعرفة والعلاقات الاجتماعية (فاركلوف٢٠٠٩). ولذا فإن هذا الاتجاه يستند إلى المؤثرات الفكرية والفلسفية التي أشرنا إليها سابقا، ويتبنى في الغالب وجهة نظر تكوينية constructive تركز على الدور الأساسي للغة في خلق الروابط والعلاقة التي تنتظم العالم وتنظمه بطريقة معينة. ومع ذلك، يميز كثير من المشتغلين في تحليل الخطاب (انظر مثلا: Fairclough 1985، و Gee 2014) بين نوعين من أنواع تحليل الخطاب ضمن هذا الاتجاه: أحدهما وصفي، وثانيهما نقدي. ومع اشتراكهما في كثير من الأسس والمنطلقات وتداخلهما في كثير من الأعمال المعنية بتحليل الخطاب على نحو يجعل الفرق بينهما فرقا في النزعة أكثر من كونه فرقا جوهريا حاسما، إلا أن الفارق الأساسي يتمثل في أن (تحليل الخطاب الوصفي) يركز على الوصف اللساني البحت لأنماط الاستعمال اللغوي من أجل تحديد وتعيين السمات والخصائص اللغوية البحتة لمجال تداولي معين (مثل الخطاب الإعلامي، والمدرسي ونحوها). وأما (التحليل النقدي للخطاب) فيحاول أن يربط التحليل اللغوي بالاعتبارات السياسية والخلفيات الأيديولوجية التي تمثل علاقات السلطة الاجتماعية وتوجه طريقة استعمال اللغة على هذا الأساس (van Dijk 1993، و Wodak & Meyer 2009).

٦. ٢. ١. التحليل النقدي للخطاب

ينتمي التحليل النقدي للخطاب إلى تيار معرفي أوسع يتبنى الرؤية النقدية في الفكر الفلسفي والبحث الاجتماعي بشكل عام (Weiss & Wodak 2003). ومن أهم ما يميز هذا الاتجاه بشكل عام هو معارضة ادعاءات الموضوعية التامة في البحث الاجتماعي، والتأكيد على أثر الانتماءات والارتباطات الاجتماعية لمؤسسات إنتاج المعرفة وصناعتها في توجيهها باتجاه المصالح الخاصة لفئات معينة، بالإضافة إلى تبني موقف أخلاقي وسياسي واضح وصريح يتمثل في العمل على كشف علاقات السلطة والإيديولوجيات المتخفية التي تحكم المجتمع وتوجه التصورات المعرفية. ولذا فإن الاتجاه النقدي يعزو إلى العلم دورا تحريريا. ومعنى ذلك ألا يقتصر دور العلم على وصف الأحداث في الواقع الاجتماعي وتفسيرها، بل ينبغي أن يسعى إلى تحرير الوعي الاجتماعي بتقديم النقد الذي يمكن الناس من الانعتاق والتخلص من أسر القيود



والأشكال المختلفة للهيمنة المعرفية. وقد توسع هذا التوجه النقدي مؤخرا، وأصبح له امتدادات متعددة في كافة مجالات البحث العلمي في الموضوعات الإنسانية، ومن ضمنها الدراسات اللغوية انظر مثلا:

(Fowler et al. 1979، Fairclough 1989 و ،Kress & Hodge 1979). والهدف الأساسي لهذا الاتجاه في الدراسات اللغوية هو تحقيق فهم نقدي للأدوار التي تؤديها اللغة إما في تشكيل المعرفة وتدويرها، أو في تنظيم المؤسسات الاجتماعية، أو في تأسيس السلطة وممارستها والمحافظة عليها.

والتحليل النقدي للخطاب امتداد لهذا الاتجاه، لأنه يهتم بالدرجة الأولى بكشف الجوانب المرتبطة بالسلطة والأيدولوجيات التي تحرك الخطاب، وتوجه استخدام اللغة بطريقة معينة، ويحاول عبر التحليل اللغوي إبراز ما ينطوي عليه الخطاب من مواقف أيديولوجية وافتراضات مضمرة تحتجب عبر استخدام خيارات لغوية معينة تجعلها تبدو طبيعية دون أن يلحظها المشاركون في الخطاب. ولذا فإن التحليل النقدي ليس مهتما بشكل أساسي بتحليل الوحدات والتصرفات اللغوية بذاتها، بل باعتبارها مكونا من مكونات السلطة والأيديولوجيا. ولا يعني ذلك بالضرورة أن التحليل النقدي يحمل موقفا سلبيا أو معارضا تجاه الخلفيات السياسية التي تشكل الخطاب (Wodak & Meyer 2009). كل ما في الأمر هو أن هذا الاتجاه يتناول الجوانب السياسية بشكل مباشر وصريح، ويعتبر أن اللغة لا يمكن إطلاقا أن تكون محايدة، لأنها دائما مندمجة في ارتباطات اجتماعية معينة، ولذا فإنها لابد أن تعكس تلك الارتباطات بدرجات مختلفة سواء بقصد أو بدون قصد من قبل مستخدميها. والتحليل النقدي للخطاب يحاول الكشف عن تلك الروابط الظاهرة والخفية بين اللغة من ناحية، والبنى السياسية والاجتماعية التي تنتمي إليها وتمثلها من ناحية أخرى، بدلا من الاكتفاء بوصف أنماط الاستعمال اللغوي وكأنها اختيارات محايدة أو مواضعات اعتباطية بحتة (van Leeuwen 2009، و Fairclough 1985). وكما يوضح الباحثون في هذا المجال (Fairclough 1985 و Gee 2014) فإن مظاهر الاستعمال اللغوي التي تعكس المنطلقات السياسية والأيديولوجية متنوعة، وتظهر في الخطاب بدرجات متفاوتة، فبعضها واضح وبارز للعيان، وبعضها أكثر غموضا وخفاء. كما أنها ترتبط بمستويات لغوية متعددة، فبعضها يرتبط بنوعية الوحدات اللغوية المستخدمة مثل المفردات



والتعابير التي تحمل دلالات مبطنة توجه الذهن والتفكير بطريقة معينة. ومنها يرتبط بطريقة تنظيم الوحدات اللغوية وتوظيفها إما عبر استخدام التشبيهات والاستعارات والكنايات لتحقيق غايات معينة، أو عبر الأساليب النحوية التي تشكل النص بطريقة معينة بالتقديم والتأخير والحذف والإضمار ونحوها. وبعضها يرتبط بالتصرفات والسلوكيات والموضوعات المطروقة أو المهملة أكثر من ارتباطه بالنظام اللغوي نفسه (van Leeuwen 2009). ويحاول التحليل النقدي استكشاف كل هذا الأساليب التي تحاول التعمية والإخفاء أو تطبيع المنطلقات والخلفيات والافتراضات الأيديولوجية الخاصة عبر إظهارها بمظهر المسلمات العامة من قبل الجميع (Fairclough 1985).

ومع أن هذا هو المنطلق الأساسي والهدف العام للتحليل النقدي للخطاب، إلا أن هذا المجال ليس له إطار نظري موحد ومحدد ومكتمل، بل هو أشبه ببرنامج بحث متغير ومتطور (van Dijk 2001). وكما يقول بعض الباحثين في هذا المجال (Wodak & Meyer 2009)، فإن من مقاصد هذا المجال أن يبقى متغيرا ومتطورا بشكل دائم ومستمر لكي يتجنب الوقوع في فخ النمطية والقولبة الأيديولوجية، وهو ما يحذر منه ويحاول الابتعاد عنه. ولذا فإن الباحثين في هذا المجال رغم اشتراكهم في الأسس والمنطلقات العامة التي شرحناها هنا، إلا أنهم ينتمون إلى مشارب متعددة ويستندون إلى نظريات مختلفة في تحليلاتهم النقدية (للاطلاع على مناقشة وافية لهذا الموضوع، انظر: (Wodak & Meyer 2009)[١].

٦. ٢. ٢. الخطاب السياسي في الإعلام

كما رأينا أعلاه، فإن التحليل النقدي للخطاب يركز بشكل أساسي على الجوانب السياسية في الخطاب ويعتبر أن السياسة حاضرة دائما في جميع المناشط الإنسانية. وبالإضافة إلى هذا المفهوم العام للسياسة، هناك مفهوم أخص يتمثل في السياسة الخالصة التي ترتبط بمؤسسات الحكومة والدولة والعلاقات الدولية وما يتعلق بذلك من مضامين وممارسات وفاعلين (Wilson 2001). والسياسة بهذا المفهوم المضيق تمثل حقلا خطابيا خاصا (Chilton 2004). وبطبيعة الحال، فإن التحليل النقدي للخطاب

١– نود أن نشير هنا إلى أن كتاب Methods for critical discourse analysis الذي أحلنا إليه قد ترجم إلى العربية وصدر عام ٢٠١٤م ضمن منشورات المركز القومي للترجمة في مصر. ولكننا رجعنا إلى النسخة الإنجليزية كما هو مذكور في الإحالات هنا وقائمة المراجع لعدم تمكننا من الحصول على النسخة العربية.



بحكم تركيزه على الجوانب السياسية بشكل عام سيكون هذا الحقل بالذات من أولويات اهتماماته البحثية (van Dijk 2001؛ Wodak & Meyer 2009). فالسياسة بهذا المعنى ذات ارتباط وثيق ومباشر بالأيدولوجيا، والسلطة، وتوجيه الرأي العام. ولذا فإن الوظائف الرئيسية للخطاب السياسي تتمثل في استمالة الجمهور، وحشد الأتباع، وتلوين الوقائع بما يخدم توطيد النفوذ والسلطة. وليس بالضرورة أن يكون ذلك واضحا وصريحا، بل كثيرا ما يلجأ الخطاب السياسي إلى استخدام استراتيجيات خفية وغير مباشرة في توظيف اللغة من أجل الإقناع والتمويه وافتعال نوع من الاستقطاب داخل المجتمع (للمزيد حول هذا الموضوع، انظر: Beard 2000، و Wilson 2001، و Chilton 2004).

وكما هو معروف فإن الخطاب السياسي، بهذا المعنى الخاص، متنوع ويتضمن العديد من الحقول والمجالات الفرعية مثل المجادلات أو الأحاديث البرلمانية، والدعاية أو الحملات الانتخابية، بالإضافة إلى الوثائق والدساتير والمعاهدات السياسية، وغيرها كثير. ومن بين هذه المجالات يظهر أن خطاب السياسة الإعلامي في وسائل الإعلام الجماهيرية هو أكثرها بروزا وأبعدها أثرا. وتمثل (الأخبار) أو (النشرات الإخبارية) أحد أهم أشكال خطاب السياسة الإعلامي بهذا المعنى. فالأخبار نوع مخصوص من الخطاب له بنية محددة، وسمات خاصة من ناحية الموضوعات والصياغة. فالوظيفة الأساسية للأخبار هي نقل الأحداث (المهمة) عالميا ومحليا عبر نصوص مكتوبة أو مسموعة تروي أهم ما يتعلق بالحدث من تفاصيل مثل: الوقائع التي يتضمنها الحدث، والمشاركين فيه، وأسبابه، وموضعه، ونتائجه العاجلة، وما قد يترتب عليه من آثار لاحقة أو مواقف مصاحبة. ويتم تقديم المحتوى المتعلق بهذه التفاصيل تبعا لصياغة معينة تمثل النمط الخاص بالأخبار في وسائل الإعلام. وبطبيعة الحال، فإن آحاد الأخبار التي يتم تقديمها وما تشتمل عليه من مضامين لا يمكن أن تكون نسخة مطابقة للواقع. بل لا بد أن تخضع بالضرورة لنوع من الانتقاء، إما بشكل مقصود ومتعمد، أو بشكل تلقائي وغير متعمد، تبعا لمعايير معينة، بعضها ظاهر، وبعضها خفي، تحدد المهم وغير المهم من الأحداث، ومواطن التركيز في كل حدث (للمزيد حول التفاصيل المتعلق بخطاب الأخبار وبنيتها (van Dijk 1988 و Bell 1991).

ورغم أن دراسة الأخبار بشكل عام ليست جديدة، إلا أن أغلب الدراسات في هذا



الموضوع تركز على الجوانب المهنية المتعلقة بطريقة جمع المعلومات وصياغتها وتنظيمها وتحريرها ونحو ذلك. وأما الدراسات التي تتبنى التحليل الناقد للخطاب في النصوص الإخبارية فقليلة جدا، ولم تبدأ إلا منذ فترة قريبة (لاستعراض بعضها، على سبيل المثال، انظر: van Dijk 2009). ورغم قلة الدراسات التي تتبنى هذا التوجه، إلا أنها توضح أن (الأخبار)، مثل أي خطاب آخر، ممتزجة بالأيديولوجيا وعلاقات السلطة، ولا بد أن تحمل نوعا من التحيز بأشكال ومظاهر مختلفة (Flower 1991 وvan Dijk 2009). ولذا فإن الهدف الأساسي للتحليل النقدي للخطاب هو الكشف عن هذه المظاهر ومواطنها عبر التحليل اللغوي للنصوص الإخبارية.

٦. ٢. ٣. استخدام المدونات في تحليل الخطاب

رغم أهمية الإسهامات المعرفية والمكانة الأكاديمية التي بدأ يحظى بها مؤخرا التحليل النقدي للخطاب[1]، إلا أنه لم يسلم من الانتقادات التي وجهت له سواء من الناحية النظرية أو المنهجية. ومن أبرز أوجه النقد في هذا الصدد أن أتباع هذا الاتجاه لا يميزون بين التحليل اللساني للخطاب وتأويل النصوص، ويعتمدون غالبا على انطباعاتهم وتأويلاتهم الذاتية بناء على مقاطع انتقائية يختارونها لأغراض محددة (انظر مثلا: Widdowson 1995 و Stubbs 1997). وقد دفع هذا النقد بعض الباحثين إلى البحث عن منهجية أكثر موثوقية في جمع البيانات وتحليلها بالمزج بين (التحليل النقدي للخطاب) و(لسانيات المدونات) (Baker et al. 2008 و Orpin 2005).

ومع أنه لا يمكن الوصول إلى الموضوعية المطلقة، إلا أن الإمكانات الحاسوبية والأساليب الإحصائية المستخدمة في لسانيات المدونات يمكن أن تجعل الخطوات المستخدمة في البحث والتحليل أقرب إلى الموضوعية. فالتعامل مع النصوص في المدونات المحوسبة والأدوات المستخدمة في تحليلها يتم بناء على معايير واضحة ومحددة وقابلة للاستعادة والتكرار. وبذلك يمكن استخراج (الكلمات المميزة) و(المتصاحبات اللفظية) وغيرها من المؤشرات الأخرى المرتبطة بـ(الخطاب) بناء على بروزها المتحقق فعلا في النصوص دون أن تكون موجهة بأفكار مسبقة لدى الباحث (Baker et al

١- على سبيل المثال هناك عدد من الدوريات العلمية المخصصة تماما أو غالبا للأبحاث المعتمدة على هذا الاتجاه، مثل: (Discourse and society) و(Critical Discourse Studies) و(the journal of language and politics)، وللمزيد حول المكانة المؤسسية للتحليل النقدي للخطاب، انظر: Wodak & Meyer, 2009.



2008). ولعل ذلك يضيق التحيز إلى أقصى حد ممكن ويجعله في حدود المقبول منهجيا، بدلا من الاعتماد على الانطباعات الذاتية والمبالغة في التعميم بناء على بيانات غير ممثلة (Baker 2012). وبناء على ذلك فإننا سنحاول في الدراسة الحالية استكشاف التوجهات التي تحكم التغطية الإخبارية من خلال الاعتماد على الكلمات الأكثر تكرارا، والكلمات المميزة، ومصاحباتها سعيا لبناء حقول دلالية تنطلق من البيانات التي تم جمعها، وليس من تصور مسبق لدى الباحثين حيال موضوع الدراسة كما سيرد تفاصيل ذلك في منهجية البحث.

٦. ٣. الدراسات السابقة

على الرغم من وجود عدد كبير من الدراسات في تحليل المحتوى الإعلامي، إلا أننا اخترنا الاقتصار على مراجعة الدراسات التي تتوفر فيها السمات التالية: (١) الدراسات الاختبارية، (٢) المكتوبة باللغة العربية فقط، (٢) والتي تعالج المضامين السياسية في وسائل الإعلام من زاوية تحليل المحتوى أو الخطاب حصرا، (٣) ضمن إطار زمني محدد، إما (٤) بالمقارنة بين وسيلتين إعلاميتين، (٥) أو بالتركيز على موضوع الدراسة الحالية (أحداث غزة). وقد حصلنا على عدد محدود من الدراسات التي تتوفر فيها هذه الشروط، وسنستعرض أهم ما ورد فيها بإيجاز فيما يلي.

بَحَثَ كل من محسن (٢٠٠٩)، والبغدادي (٢٠١٠) التغطية الصحفية لقناتي الجزيرة والعربية لأحداث الحرب السابقة على غزة في نهاية عام ٢٠٠٨م. ورغم اعتماد الدراستين على تحليل المضمون الإخباري؛ إلا أنهما تختلفان في طريقة تناول الموضوع. أما الدراسة الأولى (محسن، ٢٠٠٩) فتتكون عينتها من ٨٨ تقريرا صحفيا موزعة بالتساوي بين القناتين منذ بداية ديسمبر ٢٠٠٨ وحتى نهاية يناير ٢٠٠٩. وهدفها استخلاص مجموعة الأطر التي اعتمدت عليها القناتان في تقديم الحدث، واستكشاف دور اللغة في بناء هذه الأطر عن طريق تحليل الأساليب وأدوات التعبير المستخدمة في عرض الحدث وما يتصل به من وقائع. وباستعراض الأفكار الرئيسية في التقارير انتهت إلى تركيز الجزيرة على الزاوية الإنسانية للحدث، بينما تركز العربية على ما يجري في المكان من أحداث. وأشارت الدراسة كذلك إلى ارتفاع نسبي في الاهتمام بالسرد الوصفي والتقييمي في قناة الجزيرة، في مقابل التركيز على المعلومات في قناة العربية. وأما من ناحية أساليب التعبير، فقد أوضحت الدراسة أن الجزيرة أكثر اهتماما بتوظيف



المؤثرات اللفظية في التقارير، وأشارت كذلك إلى أن توظيف الصفات كان من أكثر الأساليب المستخدمة لبناء الإطار المستخدم في تقديم الحدث. ورغم أهمية الدراسة، إلا أنه ليس واضحا ما هي النظرية اللغوية التي اعتمدت عليها في الربط بين الأساليب اللغوية والأطر المستخدمة في تقديم الحدث.

وأما الدراسة الثانية (البغدادي ٢٠١٠) فاعتمدت كذلك على تحليل مضمون عينة من النشرات الإخبارية (٢٨ نشرة)، واللقاءات الخاصة (٨ حلقات)، وبرامج الرأي (٢٣ حلقة) المرتبطة بالحدث في القناتين. وأداة التحليل المستخدمة عبارة عن استمارات خاصة أعدتها الباحثة لهذا الغرض. وانتهت الدراسة إلى التشابه بين القناتين في الزمن المخصص لتغطية الحدث، والاعتماد على المصادر الذاتية في نقل الخبر، بالإضافة إلى التوزان في محتوى الخبر من ناحية الاهتمام بتقديم كافة وجهات النظر المختلفة حول الحدث. ولكن الاختلاف الرئيسي يتمثل في نوعية المصطلحات المستخدمة في توصيف بعض جوانب الحدث. فالجزيرة تستخدم كلمات من قبيل «الحرب، المقاومين، الشهداء» في مقابل استخدام العربية لـ «الهجوم، المقاتلين، القتلى». ومع أن الباحثة لم تقدم إحصاءات عن نسبة استخدام هذه المفردات في النصوص التي قامت بتحليلها، إلا أنها أشارت إلى مقابلات أجرتها مع المسؤولين عن التحرير في القناتين توضح أن استخدام هذه الكلمات مقصود بناء على السياسة المهنية التي تنتهجها كل منهما بمبررات وذرائع معينة.

وأما في سياق الدراسات المقارنة للتغطية الصحفية لموضوعات أخرى متعلقة بالقضية الفلسطينية فقد وقفنا على دراستين أيضا (بصيص ٢٠٠٧؛ أبو مزيد ٢٠١٣). ففي الدراسة الأولى (بصيص ٢٠٠٧) قام الباحث برصد الخطاب الصحفي لثلاث صحف جزائرية مختلفة: إحداها حكومية (الشعب)، والأخرى خاصة (الخبر)، والثالثة مستقلة (الحرية)، حيال انتفاضة الأقصى بمقارنة كافة ما ورد فيها من نصوص عن هذا الموضوع خلال الفترة من ٢٨/ ٩/ ٢٠٠٠ إلى ٣٠/ ٤/ ٢٠٠٣. واعتمد الباحث على تحليل المضمون من أجل الكشف عن الإطار الإعلامي الذي يحكم التوجهات السياسية لكل صحيفة تجاه الموضوع. وبينت الدراسة نتائج عدة، من أبرزها: عدم اتساق المعالجة الصحفية للصحف الثلاث، واختلافها بشكل واضح حيال القضايا الرئيسة في سياق القضية الفلسطينية (القدس، اللاجئين، حق العودة، الاستيطان).



كما أظهر تحليل الخطاب اختلاف الأطر المرجعية في النظر إلى الموضوع، حيث ركزت صحيفتا «الشعب» و «الخبر» على الأطر العربية والإسلامية للقضية، خلافا لصحيفة «الحرية» التي ركزت على الأطر المحلية والدولية.

وأما الدراسة الثانية (أبو مزيد ٢٠١٣) فهي دراسة مقارنة لتناول صحيفتين فلسطينيتين (الحياة الجديدة، فلسطين) لموضوع المصالحة بين منظمة التحرير الفلسطينية وحركة حماس خلال الفترة من فبراير ٢٠٠٩ م وحتى فبراير ٢٠١٢ م. واعتمدت الدراسة على عينة عشوائية من أعداد الصحف التي صدرت أثناء الأحداث الرئيسية في مسار المصالحة. وأما من ناحية منهجية التحليل، فقد اعتمدت الدراسة على تحليل المضمون بناء على أدوات أعدتها الباحثة بنفسها من أجل هذا الغرض. وانتهت الدراسة إلى وجود تباين واضح بين الصحيفتين في تناول الموضوع بسبب تأثير التوجه الأيديولوجي والارتباط المؤسسي (أو الحزبي) لكل منهما، حيث قامت كل من الصحيفتين بإسناد الصفات والأدوار الإيجابية للجهة التي تنتمي إليها، وإسناد الصفات والأدوار السلبية للجهة المقابلة.

وأما في سياق المقارنة العامة بين قناتين إعلاميتين في موضوع معين، فقد وقفنا أيضا على دراستين (الخطيب ٢٠١١، وعبدالغني ٢٠١١). أما الدراسة الأولى (الخطيب ٢٠١١) فتقارن التحليل الإعلامي بين قناتي الجزيرة والحرة لسياسات الرئيس باراك أوباما داخليا وخارجيا. وتتمثل عينة الدراسة بحلقات برنامجي (من واشنطن) في قناة الجزيرة، و(داخل واشنطن) في قناة الحرة على مدى ثلاثة أشهر خلال عام ٢٠٠٩م. واعتمدت الدراسة أيضا على منهجية تحليل المضمون بناء على أداة أعدت لهذا الغرض. وانتهت الدراسة إلى وجود تباين كبير بين القناتين في نوعية الكلمات الدالة رغم معالجة الموضوعات نفسها. ومن ناحية أخرى، فقد غلب على قناة الجزيرة التركيز على انعكاس القضايا المطروحة على الواقع العربي والإسلامي، بينما كانت قناة الحرة تركز على انعكاس تلك القضايا على الداخل الأمريكي.

وأما الدراسة الثانية (عبدالغني ٢٠١١) فكان هدفها المقارنة بين تناول القضايا العربية في قناتي الجزيرة الإنجليزية والبي بي سي الدولية. وتتمثل عينة الدراسة في ٦٠ ساعة إخبارية على مدى شهر كامل من نشرة منتصف اليوم في القناتين، موزعة



بالتساوي بينهما، بالإضافة إلى نماذج من البرامج الإخبارية، بلغ عددها ٢٢ حلقة من كل قناة. واعتمدت الدراسة أيضا على تحليل المضمون بناء على أداة أُعدت لهذا الغرض. وقد انتهت الدراسة إلى وجود فرق في الملامح اللغوية بين القناتين، خصوصا في تناول القضايا ذات الصلة بالشأن العربي، بينما كان الاختلاف أقل في القضايا غير العربية. وعلى سبيل المثال، تستخدم البي بي سي التعبيرات التالية: «العنف الإسلامي»، «أمن إسرائيل»، «سفن مساعدات» في مقابل استخدام الجزيرة لتعبيرات أخرى من قبيل: «هجمات في عواصم أوروبية»، «تصفية المقاومة»، و»حصار غزة» في الإشارة إلى الموضوعات نفسها. وأظهرت الدراسة كذلك اختلاف الاستراتيجيات الجدلية التي تتبناها كل قناة.

ورغم أن حجم الدراسات التي تم استعراضها هنا قليل نسبيا، ولا يساعدنا على استخلاص نتائج واضحة؛ إلا أن هذه الدراسات تشير إلى حد ما إلى أن وسائل الإعلام تتباين في التحليل، والمضمون اللغوي، وأطر التناول، رغم تناولها للموضوعات نفسها. وبطبيعة الحال، هناك أسباب كثيرة لهذا التباين. ولكن يظهر أن حجم التباين والاختلاف في الصحف الحزبية والمحلية أكثر بكثير منه في الصحف العالمية والقنوات الدولية ذات الانتشار الواسع. فالصحف الحزبية الصغيرة والمحلية عادة تحمل وجهات نظر ضيقة تعكس رؤية المؤسسات التي تنتمي إليها بشكل مباشر. وهذه النتيجة تتفق بشكل عام مع مقارنات مماثلة أجريت على الصحف البريطانية (انظر مثلا: Baker 2012). ويأتي في الدرجة الثانية الاختلاف بين القنوات العربية (مثل الجزيرة)، والقنوات الأخرى ذات المرجعيات الغربية (مثل: الحرة والبي بي سي). وأما القنوات العربية (الجزيرة والعربية) التي تمثل الموضوع الأساسي للدراسة الحالية، فرغم أن تغطيتها لأحداث غزة متشابهة في كثير من الجوانب (انظر: البغدادي، ٢٠١٠)، إلا أنه يوجد بينها أيضا عدد من الاختلافات. وأكثر ما يتضح الاختلاف بينهما في الجوانب المتعلقة بأطر الرؤية، وأساليب السرد (محسن ٢٠٠٩)، بالإضافة إلى نوعية الكلمات المستخدمة في توصيف الحدث (البغدادي ٢٠١٠).

ومع أهمية هذه النتائج التي استخلصناها من الدراسات السابقة، إلا أنه يلاحظ أن كل الدراسات التي تم استعراضها هنا قام بإعدادها متخصصون في الإعلام ذوو



مرجعية إعلامية، وليسوا لغويين متخصصين في التحليل اللغوي (وللاطلاع على بعض الإشكاليات المتصلة بذلك في الدراسات العربية، انظر: شومان ٢٠١١). ولذا فإنها لا تعتمد على نظريات لغوية محددة في التحليل اللغوي. وتبعا لذلك، فإنها لا تنطلق من نظرية «التحليل النقدي للخطاب» كما استعرضناها في الإطار النظري للدراسة الحالية. وبالإضافة إلى ذلك، فإن أدوات التحليل المستخدمة في هذه الدراسات أعدها الباحثون بأنفسهم، ولم تعتمد على منهجيات التحليل المستخدمة في (لسانيات المدونات) والتي تنطلق من البيانات نفسها. ولذا فإنه يطالها النقد الذي ذكرناه سابقا عن الانتقائية والتأويلات الذاتية في تحليل الخطاب.

٦. ٤. وصف بيانات الدراسة

تركز الدراسة الحالية، كما ذكرنا سابقا، على التحليل النقدي للخطاب في النصوص الإخبارية المتعلقة بحرب غزة (عام ٢٠١٤م) في قناتي الجزيرة والعربية. وقد تم اختيار هاتين القناتين بالتحديد لثقلهما الكبير في العالم العربي من حيث التأثير وعدد المتابعين. وتتمثل عينة الدراسة على وجه التحديد بالنصوص الإخبارية لمدة ٥١ يوما منذ بداية الحرب بتاريخ ٧/ ٧/ ٢٠١٤م وحتى نهايتها بالهدنة في تاريخ ٢٧ / ٨ / ٢٠١٤م. وقد جمع الباحثان المدونة بالاعتماد على الموقع الشبكي الخاص بكل قناة على الإنترنت، من خلال الاقتصار على نصوص الأخبار، واستبعاد النصوص الأخرى كالحوارات، وبرامج التحليل السياسي، وتعليقات القراء، وغيرها. ومن خلال استثمار خاصية البحث في موقعي القناتين تم الوصول لكافة النصوص الإخبارية. وبعد جمع النصوص تم تحويها إلى صيغة نصية (text)، وتنقية الملفات من الشوائب مثل الصور وبعض التعليقات على الأخبار والدعايات ونحوها. وبعد ذلك تم جمع المستندات النصية التابعة لكل قناة في ملف مستقل، موزعة تبعا لتسلسها التاريخي بحيث تم جمع أخبار كل يوم على حدة.

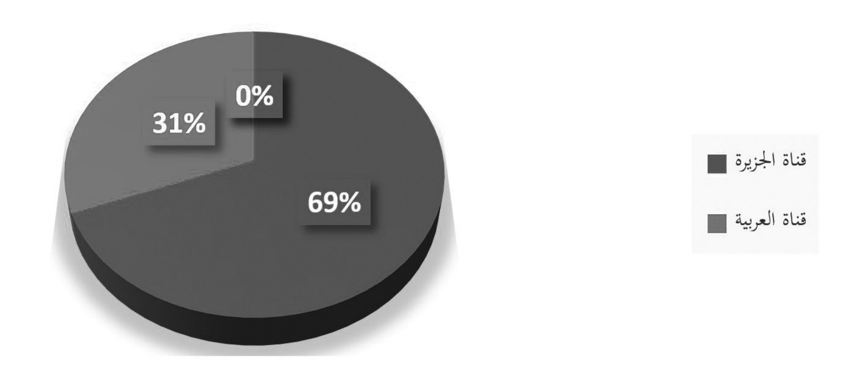
وقد بلغت حصيلة المدونة اللغوية بشكل عام ٥٢٨٦٤١كلمة، موزعة على ١٥٢٥خبرا (نصا)، كما يوضح الجدول (١).



المصدر	عدد النصوص	عدد الكلمات الكلي	عدد الكلمات دون تكرار	متوسط طول النص
الجزيرة	١٠٥٤	٤٠٨٢٧٤	٥٢٩٤٠	٣٨٧
العربية	٤٧١	١٢٠٣٦٧	٢٥٢٢٥	٢٥٦
كامل المدونة	١٥٢٥	٥٢٨٦٤١	٧٨١٦٥	٣٤٧

الجدول (١): عدد نصوص المدونة، وكلماتها، ومتوسط الطول للنصوص.

وكما يلاحظ في الجدول السابق، يوجد تفاوت بين القناتين في عدد الأخبار وطولها. فمن حيث العدد بلغت أخبار الجزيرة قرابة ٦٩٪ من نصوص المدونة، بينما بلغت أخبار العربية ما نسبته تقريبا ٣١٪ من المدونة، كما يوضح الشكل (١). ومن ناحية أخرى، تتسم أخبار الجزيرة بالطول نسبيا مقارنة بأخبار العربية حيث بلغ متوسط الخبر في الجزيرة (٣٨٧ كلمة)، بينما بلغ متوسط عدد كلمات الخبر في قناة العربية (٢٥٦ كلمة).



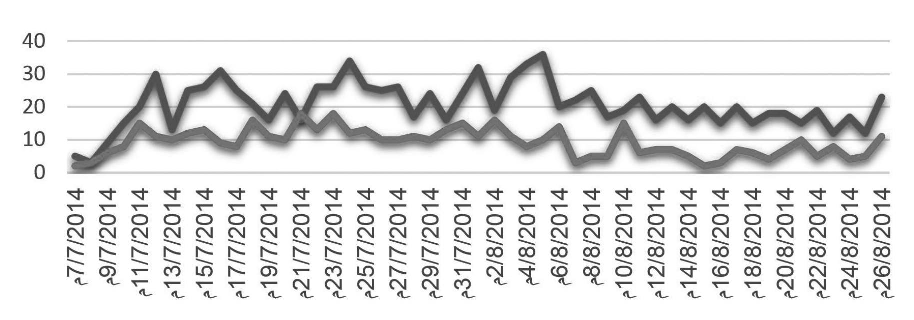
الشكل (١): نسبة النصوص في المدونة مقسمة على القناتين.

وأما من ناحية توزيع الأخبار على الأيام في التغطية الإعلامية للقناتين، فأظهرت البيانات وجود تفاوت في عدد الأخبار من يوم إلى آخر، كما يوضح ذلك الشكل (٢).



ويرجع ذلك إلى طبيعة الأحداث، حيث تشهد بعض الأيام كثافة في الأخبار نتيجة لتعدد الأحداث وأهميتها، بينما شهدت بعض الأيام محدودية في الأخبار. ويلاحظ تشابه القناتين في ارتفاع وانخفاض عدد الأخبار في الأيام المختلفة، مع زيادة أخبار الجزيرة على وجه العموم.

عدد الأخبار اليومية من قناتي الجزيرة والعربية خلال فترة حرب غزة



الشكل (٢): عدد الأخبار في قناتي الجزيرة والعربية موزعة على الأيام.

وتجدر الإشارة هنا إلى اختلاف القناتين في طريقة تحديث الأخبار. فقناة العربية تقوم بإدراج التحديثات على ذات الخبر، وفي نفس الصفحة، وتستمر التحديثات لليوم التالي، ويشار إلى ذلك في صفحة الخبر (مثلا: تاريخ الخبر: ١٠/ ٨/ ٢٠١٤، تاريخ آخر تحديث ١١/ ٨/ ٢٠١٤)؛ فيكون تاريخ الخبر، وتاريخ تحديث الخبر مختلفين في كثير من الأخبار التي شهدت زخما إعلاميا. أما قناة الجزيرة فتكتفي غالبا بتحديث الخبر في نفس اليوم. وفي حال ورود أخبار جديدة في اليوم التالي يتم إدراجها على هيئة خبر جديد، وكان لذلك أثره في عدد الأخبار، وفي توزعها على المدى الزمني للأحداث.

٦. ٥. تحليل المدونة والأدوات المستخدمة في التحليل

بعد جمع نصوص المدونة؛ تم ترتيبها، وتوحيد طريقة تسمية ملفاتها – بحيث يتكون اسم الملف من المصدر (الجزيرة) أو (العربية)، ثم تاريخ النص، ثم ترتيبه في ذلك التاريخ–، ثم وضعت النصوص التي تنتمي إلى كل قناة في مجلد منفصل. وبعد ذلك قمنا بالخطوات الثلاث التالية بالتتابع لتجهيز المدونة للتحليل:



الأولى: التشذيب باستخدام برنامج المشذب العربي[١]. ويشتمل التشذيب على: (أ) فصل علامات الترقيم والأرقام الملتصقة بالكلمات، (٢) وإزالة علامات التطويل (الكشيدة) الموجودة أواسط الكلمات، (٣) وإزالة الفراغات الزائدة بين الأسطر والكلمات. والهدف من ذلك تقليل تشتت البيانات.

الثانية: التجزئة باستخدام برنامج المجزئ العربي[٢]. وتتمثل التجزئة في فصل حروف الجر (اللام والباء)، والعطف (الواو والفاء)، والضمائر المتصلة، ونحوها من السوابق واللواحق الأخرى التي تلتصق بالكلمات. والهدف من هذه الخطوة هو استخراج الكلمات المتماثلة دون التأثر بنوعية السوابق واللواحق التي تتصل بها.

الثالثة: التحليل الكمي باستخدام الإصدار الأخير (٤،٥) من برنامج معالجة المدونات العربية (غواص)[٣]. ويتضمن برنامج غواص عدد من الوظائف التي تمكن المستخدم من تحليل المدونات العربية؛ ومنها: (أ) استعراض قوائم التكرار والتكرار النسبي، إما لكل كلمات المدونة أو لبعضها بحسب احتياج المستخدم، (ب) استعراض الكشاف السياقي لأي كلمة يحددها المستخدم، من خلال عرض السياق الذي وردت فيه الكلمة بحسب العدد الذي يريده المستخدم، (ج) استخراج الكلمات المميزة لأي مدونة مقارنة بمدونة أخرى بالاعتماد على عدد متنوع من المعايير الإحصائية (يوفر البرنامج ٨ معايير مختلفة و٣ محددات مختلفة لتصفية النتائج)، (د) استخراج المتصاحبات أو المتلازمات اللفظية في المدونة بالاعتماد على عدد متنوع من المعايير الإحصائية (يوفر البرنامج ٩ معايير إحصائية، و٣ محددات مختلفة لتصفية النتائج)[٤].

وفي الدراسة الحالية قام الباحثان بالتعرف على التوزيع الإحصائي للكلمات في

١- تم تصميم برنامج المشذب العربي، وبقية الأدوات الأخرى التي استخدمناها للتعامل مع المدونة وتحليلها في الدراسة الحالية، في مدينة الملك عبدالعزيز للعلوم والتقنية. ويمكن تحميل المشذب العربي من الرابط التالي: http://sourceforge.net/projects/ghawwasv4/files/?source=navbar.

٢- ويمكن الحصول على معلومات أكثر حول مجزئ ستانفورد العربي من خلال الرابط التالي: http://nlp.stanford.edu/software/segmenter.shtml.

٣- ويمكن الحصول على البرنامج من الرابط التالي: http://sourceforge.net/projects/ghawwasv4/files/?source=navbar.

٤- يمكن الحصول على معلومات أكثر تفصيلا حول برنامج (غواص) من خلال دليل المستخدم المضمن مع البرنامج في الرابط المذكور في الهامش السابق، وللاطلاع على نماذج ممثلة لاستخدام (غواص) في تحليل المدونات العربية، يمكن الرجوع إلى: الشمري والثبيتي، ٢٠١٥.



مدونتي الجزيرة والعربية، واستخراج الكلمات المئة الأكثر تكرارا في كل منهما (سيتم استعراض النتائج في الجزء التالي). وبالإضافة إلى استخراج الكلمات الأكثر تكرارا، قام الباحثان باستخراج الكلمات المميزة لكل مدونة. ويتم التعرف على الكلمات المميزة من خلال مقارنة التوزيع الإحصائي للكلمات في المدونة المستهدفة بالتوزيع الإحصائي للكلمات في مدونة مرجعية أخرى. فالكلمات المميزة هي الكلمات التي يختلف توزيعها الإحصائي في المدونة الرئيسية (المقصودة) بشكل كبير عن توزيعها الإحصائي في المدونة المرجعية التي تتم مقارنتها بها.

وقام الباحثان باستخراج الكشاف السياقي للكلمات المهمة للتحقق من تفضيلها الدلالي وللاطلاع على السياق الذي وردت فيه الكلمات. وبعد ذلك قام الباحثان باستخراج المصاحبات اللفظية للكلمات المهمة في المدونتين كلتيهما عن طريق استخدام معامل الاحتمال اللوغاريثمي أيضا، عند نسبة خطأ أقل من واحد في المليون ($p <$ ٠,٠٠٠٠٠١)، أي أن قيمة معامل الاحتمال اللوغاريثمي للكلمة يجب أن تساوي أو تكون أكبر من ٤٠ (الجدول ٢ و٣). وبالإضافة إلى ذلك، تم اختيار المحددات التالية لتصفية النتائج: (١) أن يكون التكرار الكلي للكلمة المصاحبة أكبر من أو يساوي ٢٠٪ من التكرار الكلي للكلمة الرئيسية، (٢) أن يكون تكرار الكلمة المصاحبة في السياق أكبر من أو يساوي ٣٠٪ من تكرارها الكلي في المدونة، وسنستعرض أهم نتائج البحث وفقا للقيم السابقة في الجزء التالي.

ومع وجود طرق عديدة لاستخراج الكلمات المميزة؛ إلا أن الباحثان استخدما معامل الاحتمال اللوغاريثمي كمقياس إحصائي عند نسبة خطأ أقل من واحد في الألف ($p <$ ٠,٠٠١). أي أن قيمة معامل الاحتمال اللوغاريثمي للكلمة يجب أن تساوي أو تكون أكبر من ٩,٠٠ (الجداول ٥، و٨، و٩، و١٠). وقد اختير معامل الاحتمال اللوغاريثمي لأنه لا يفترض أي توزيع إحصائي للكلمات في المدونة، ويمكن استخدامه لتحليل المدونات ذات الحجم الصغير والكبير على حد سواء. وبعد حصر الكلمات التي تحقق الشرط الإحصائي السابق باستخدام معامل الاحتمال اللوغاريثمي، قام الباحثان باختيار المحددات التالية لاستخراج الكلمات المميزة: (١) أن يكون تكرار الكلمة في المدونة الرئيسية أكبر من أو يساوي عشر مرات، و(٢) أن تكون نسبة ورود



الكلمة في المدونة الرئيسية أكثر من نسبة ورودها في المدونة المرجعية بخمسة أضعاف. وبهذه الطريقة استطاع الباحثان تحديد الكلمات التي تتميز بها النصوص الإخبارية في قناة الجزيرة مقارنة بنصوص الأخبار في قناة العربية، والكلمات التي تتميز بها الأخبار في قناة العربية مقارنة بالأخبار في قناة الجزيرة.

وبعد استخراج الكلمات الأكثر تكرارا، والكلمات المميزة في المدونتين، قام الباحثان بتصنيفها دلاليا إلى أبرز المحاور التي تنتمي لها. فمن خلال الانطلاق من البيانات التي تم جمعها ومحاولة تصنيفها دلاليا؛ تم بناء محاور عدة تنتمي إليها قائمة الكلمات، مثل: الكلمات التي تعبر عن الأطراف المشتركة في الحدث، ومكانه، والمعنيين به، والكلمات التي تصف الحدث بشكل مباشر، والكلمات التي تشير إلى مواقف مصاحبة للحدث ومرتبطة به، والكلمات التي تعبر عن نوعية التغطية الصحفية للحدث. وتلك الآلية في الانطلاق من البيانات البحثية واستقرائها لبناء محاور انتمائها دلاليا تعطي التحليل البحثي عمق أكبر، وموضوعية في تحليل البيانات واستيعاب مضامينها مقارنة بالطريقة التقليدية التي تضع الحقول الدلالية ابتداء بناء على أدبيات الدراسة (Gabrielatos & Baker 2008).

٦. ٦. نتائج البحث

كما تمت الإشارة في القسم السابق، فإن أول خطوة قام بها الباحثان، بعد تشذيب المدونتين وتجزئتهما، هي استخراج الكلمات الأكثر تكرارا في كل مدونة. ويعرض الجدولان (٢) و(٣) الكلمات الأكثر تكرارا في مدونتي الجزيرة والعربية على التوالي.

وبالنظر إلى الجدولين السابقين، يُلحظ أمران. الأول: ارتفاع نسبة التوزيع الإحصائي للكلمات الأكثر تكرارا بشكل كبير جدا (تصل إلى النصف تقريبا) بالنسبة إلى بقية كلمات المدونة. أما الثاني: وجود عدد كبير نسبيا من (كلمات المحتوى) ضمن الكلمات الأكثر تكرارا.

وهذه النسبة-سواء من ناحية التوزيع الإحصائي العام، أو من ناحية ظهور (كلمات المحتوى)- تُعدّ أكثر مما هو معتاد (قارن مثلا بـ: الشمري والثبيتي ٢٠١٥ حيث بلغت نسبة مجموع التوزيع الإحصائي للكلمات الأكثر تكرارا ٢٤٪ تقريبا، والغالبية العظمى فيها-إلا ما ندر-من الكلمات الوظيفية).



#	الكلمة					٪
١	و	في	ب	ل	ان	٢٦,٥٥
٢	الى	ما	هم	اسرائيل	عن	٥,٩٦
٣	حماس	لا	قال	العدوان	الاسرائيلية	٣,١٢
٤	ف	الجزيرة	بعد	الفلسطيني	ذلك	٢,٤٥
٥	حركة	القطاع	اليوم	س	كان	١,٩٣
٦	كل	النار	الحرب	اكثر	منذ	١,٥٣
٧	او	نا	غير	قبل	اي	١,٣٠
٨	علي	عدد	القصف	اضاف	نت	١,١٦
٩	ابو	الشعب	اسم	حيث	ضد	٠,٩٩
١٠	كتائب	المصرية	تل	بعض	دون	٠,٩٢
المجموع			٤٥,٩١			

#	الكلمة					٪
١	من	ها	ه	على	غزة	٢٦,٥٥
٢	الاسرائيلي	التي	قطاع	مع	المقاومة	٥,٩٦
٣	الذي	ك	الاحتلال	الفلسطينية	بين	٣,١٢
٤	قد	اطلاق	وقف	هذه	لم	٢,٤٥
٥	هذا	هو	الفلسطينيين	خلال	الجيش	١,٩٣

#	الكلمة					٪
٦	لكن	كانت	قالت	رئيس	قصف	١,٥٣
٧	القاهرة	المتحدة	الاسلامية	مدينة	مصر	١,٣٠
٨	حتى	القسام	هي	اسرائيلية	نحو	١,١٦
٩	المدنيين	امس	صواريخ	اكد	الصواريخ	٠,٩٩
١٠	العسكرية	اخرى	قوات	ابيب	الذين	٠,٩٢
المجموع			٤٥,٩١			

الجدول (٢) الكلمات الأكثر تكرارا في مدونة الجزيرة.

#	الكلمة					٪
١	و	في	ل	ب	من	٢٦,١٢
٢	الى	اسرائيل	ما	قطاع	التي	٦,٢٥
٣	قال	اطلاق	وقف	بين	الاسرائيلية	٣,٢٦
٤	س	لا	النار	الفلسطيني	هذه	٢,٤٤
٥	ذلك	منذ	نا	لم	ف	٢,٠٣
٦	كان	قتل	هو	كل	رئيس	١,٥٦
٧	المصرية	العربية	علي	قد	عدد	١,٣٩
٨	المدنيين	اضاف	قالت	الصواريخ	قصف	١,٢٣
٩	وزارة	العام	الهجوم	اعلنت	اسرائيلية	١,١٤
١٠	حتى	مدينة	هي	فيما	صواريخ	١,٠٩
المجموع			٤٦,٥١			

#	الكلمة					٪
١	ان	ها	على	ه	غزة	٢٦,١٢
٢	الإسرائيلي	عن	هم	حماس	مع	٦,٢٥
٣	الذي	بعد	ك	الجيش	اليوم	٣,٢٦
٤	خلال	مصر	القطاع	هذا	الفلسطينية	٢,٤٤
٥	حركة	الفلسطينيين	المتحدة	اسم	القاهرة	٢,٠٣
٦	اكثر	الامم	كانت	اي	لكن	١,٥٦
٧	اعلن	قبل	الخارجية	غير	الرئيس	١,٣٩
٨	حسب	اكد	الحرب	مجلس	بيان	١,٢٣
٩	الشعب	العسكرية	حيث	المبادرة	تم	١,١٤
١٠	الامن	ضد	الثلاثاء	التوصل	او	١,٠٩
المجموع			٤٦,٥١			

الجدول (٣) الكلمات الأكثر تكرار في مدونة العربية.

ولكن ظهور هذه النسبة بهذا الشكل أمر متوقع، نظرا إلى أن المدونة المستخدمة في الدراسة الحالية صغيرة الحجم نسبيا، ومتخصصة في مجال محدد، فمن الطبيعي أن يزداد تكرار نوعية معينة من الكلمات ذات الارتباط الخاص بالموضوع.

وبالتركيز على (كلمات المحتوى) تحديدا، فإننا سنجد أن القناتين تشتركان في العدد الأكبر من الكلمات الأكثر تكرارا في الجدولين السابقين، وتنفرد كل منها بظهور عدد معين من الكلمات ضمن الأكثر تكرارا، كما يوضح الجدول (٤).



مشتركة بين القناتين	تكثر في الجزيرة	تكثر في العربية
قطاع (القطاع)/ غزة/ إسرائيل (الإسرائيلي/ الإسرائيلية/ إسرائيلية) / الفلسطيني (الفلسطينية/ الفلسطينيين) / حركة/ حماس/ مصر (المصرية)/ القاهرة/ المتحدة/ رئيس/ الحرب/ قصف/ وقف/ اطلاق/ النار/ الصواريخ (صواريخ)/ الجيش/ العسكرية/ المدنيين/ الشعب/ مدينة/ قالت (أضاف/ أكد).	المقاومة/ العدوان/ الاحتلال/ القصف/ كتائب/ القسام/ الإسلامية/ قوات/ تل (أبيب).	الأمم/ الخارجية/ الرئيس/ أعلن (أعلنت)/ مجلس/ الأمن/ بيان/ الهجوم/ قتل/ المبادرة/ التوصل

الجدول (٤) توزيع الكلمات الأكثر تكرارا بين القناتين.

وهنا لا بد من التوضيح أن ورود الكلمة ضمن الكلمات الأكثر تكرارا مرتبط بتوزّعها النسبي في المدونة، وليس بعدد مرات التكرار، وهذا مهم بسبب تفاوت عدد الكلمات في المدونتين كما ذكرنا في الجدول (١) سابقا. ويوضح ذلك الجدول (٥) الذي يحتوي على تكرار كلمة (قتل)، وتكرارها النسبي في المدونتين، فعدد تكرار (قتل) في الجزيرة أكثر من تكرارها في العربية غير أن نسبة التوزيع التكراري النسبي لمجمل مدونة الجزيرة قليل جدًا ولا يشكل قيمةً نسبية مقارنة بقيمة نسبها العالية في الجزيرة بواقع ١٢٪ تقريبا من مجمل مدونة العربية.

الكلمة	التكرار في الجزيرة	التكرار في العربية	التكرار في القناتين	التكرار النسبي في الجزيرة	التكرار النسبي في العربية	التكرار النسبي في القناتين
قتل	٣٦٦	٢٣٢	٥٩٨	٠,٠٧٣٨٦٧	٠,١٦٥٣٤٥	٠,١١٩٦٠٦

الجدول (٥): التكرار والتكرار النسبي لكلمة (قتل) في القناتين.



وبالرجوع إلى قائمة الكلمات في الجدول (٤)، نجد أن الكلمات الأكثر تكرارا في القناتين تحمل دلالات معينة، ويمكن تصنيفها تبعا لذلك إلى عدد من الحقول أو المحاور. وكما يوضح الجدول (٦)، فقد قام الباحثان من خلال استقراء البيانات وبالاعتماد على عينات عشوائية من الكشاف السياقي للكلمات، بتصنيفها إلى أربعة محاور رئيسية.

المحور الذي تنتمي إليه الكلمات		الكلمات الأكثر تكرارا		
		مشتركة بين القناتين	تكثر في الجزيرة	تكثر في العربية
أطراف الصراع/ وأماكنه/ والمعنيون به	أقاليم	القطاع/ قطاع/ غزة	————	————
	دول (أو كلمات منسوبة لها)	إسرائيل (الإسرائيلي/ الإسرائيلية/ إسرائيلية) الفلسطيني (الفلسطينية/ الفلسطينيين) مصر (المصرية)	————	————
	مدن	القاهرة	تل أبيب	————
	منظمات	حركة حماس	كتائب/ القسام/ الإسلامية	الأمم (المتحدة)/ مجلس الأمن
	قطاعات وفئات اجتماعية مرتبطة بالحدث	الجيش/ المدنيين/ الشعب	قوات	————
	كلمات تشير إلى مواضع غير محددة	مدينة	————	————



المحور الذي تنتمي إليه الكلمات	الكلمات الأكثر تكرارا		
	مشتركة بين القناتين	تكثر في الجزيرة	تكثر في العربية
كلمات تشير إلى وصف مباشر للحدث	قصف/ الحرب/ العسكرية/ وقف/ إطلاق/ النار/ صواريخ (الصواريخ)	المقاومة/ العدوان/ الاحتلال/ القصف	قتل/ الهجوم
كلمات تشير إلى مواقف مصاحبة للحدث ومرتبطة به	ــــــــ	ــــــــ	المبادرة
كلمات تشير إلى التغطية الصحفية للحدث	قالت/ أكد/ أضاف	ــــــــ	أعلن/ أعلنت

الجدول (٦): المحاور التي تنتمي لها الكلمات الأكثر تكرارا في القناتين.

وكما يوضح الجدول (٦) أعلاه، فإن القناتين متشابهتان في التركيز على عدد من الكلمات التي تمثل محاور رئيسية في تناول الحدث. فغزة وما يرتبط بها من كلمات كـ (القطاع، قطاع) هي مركز الحدث. وإسرائيل (وما يتصرف عنها من كلمات)، ومتصرفات «فلسطين» (ويلاحظ غياب اسم العلم «فلسطين»)، بالإضافة إلى (حركة حماس) تعبر عن الأطراف الرئيسية في الحدث. وتشترك القناتان كذلك في بروز كلمات (مصر، المصرية، القاهرة)، وهو ما يعبر عن التأكيد على دور مصر باعتبارها أحد الأطراف المعنية بشكل مباشر بالموضوع. وفي الوصف المباشر للحدث، تشترك القناتان في الاستخدام المكثف لكلمات (القصف، الحرب، وقف، إطلاق، النار، الصواريخ). وأما أكثر الكلمات تكرارا في القناتين معا فيما يتعلق بالقطاعات والفئات الاجتماعية المرتبطة بالحدث فهي (الجيش، المدنيين).

وفي مقابل هذه التشابهات الكثيرة، توجد بعض الاختلافات. ففي الجزيرة تكثر الإشارة بشكل مباشر إلى (كتائب القسام) و(تل أبيب)، ويكثر كذلك استخدام كلمات (المقاومة، العدوان، الاحتلال) في توصيف بعض الجوانب المرتبطة بالحدث. وأما في



العربية فتكثر الإشارة إلى (الأمم المتحدة، ومجلس الأمن)، كما أن الكلمات الأكثر تكرارا في توصيف الحدث في العربية هي (قتل، الهجوم). وفي المواقف المصاحبة للحدث، تبرز كلمة (المبادرة) باعتبارها أحد الكلمات الرئيسية في تغطية العربية.

ومن أجل الاقتراب أكثر من دلالات الكلمات الأكثر تكرارا، قمنا باستخراج المصاحبات اللفظية لها في كل من المدونتين، بحسب المعايير الإحصائية التي شرحناها سابقا. ويعرض الجدول (٧) المصاحبات اللفظية لبعض الكلمات الأكثر تكرارا في المدونتين معا.

الكلمة	المصاحبات اللفظية		
	تشترك بها القناتان	تكثر في الجزيرة	تكثر في العربية
القطاع	على	سكان/ قصف/ في/ العدوان	منذ/ شمال/ الإسرائيلي/ الهجوم
غزة	قطاع/ على/ في/ الإسرائيلي/ إسرائيل	العدوان/ غزة/ الإسرائيلية	ــــــــ
الإسرائيلي	الجيش/ العدوان/ غزة/ قطاع/ على/ القصف/ رئيس/ منذ	الاحتلال/ إسرائيل/ اليوم	الوزراء/ نتنياهو/ أعلن/ القطاع/ بدء/ الفلسطيني
الإسرائيلية	الحكومة/ الحرب/ على/ قطاع	غزة	العملية/ العسكرية
الفلسطينية	ــــــــ	المقاومة	الفصائل/ الصحة/ وزارة/ عباس
الفلسطينيين	الإسرائيليين/ بين	وكالة/ المدنيين	القتلى/ الاونروا
مصر	ــــــــ	معبر/ رفح/ مع/ حماس	ــــــــ



الكلمة	المصاحبات اللفظية		
	تشترك بها القناتان	تكثر في الجزيرة	تكثر في العربية
المصرية	المبادرة/ الخارجية	السلطات/ وقف/ رفض/ إطلاق/ النار/ وزارة/ التهدئة/ حماس	دعم
القاهرة	الوفد/ المفاوضات/ الفلسطيني/ مفاوضات/ إلى/ غير	وقف/ اتفاق/ في/ التوصل/ التهدئة/ النار/ من/ إطلاق	ــــــــــ
الحرب	غزة/ الإسرائيلية/ على	التي/ هذه	ــــــــــ
قصف	إسرائيلي/ فلسطينيين/ منزل	استشهد/ استشهاد/ مدينة/ القطاع/ ثلاثة/ رفح/ مراسل/ غزة/ اليوم	مدرسة/ قتل/ شرق/ الاونروا
وقف	إطلاق/ النار/ التوصل/ ل/ اتفاق/ المبادرة/ المصرية/ العدوان/ إلى	القاهرة/ غزة/ قطاع	أجل
الصواريخ	إطلاق/ أطلقت/ القبة/ الحديدية	المقاومة/ الفلسطينية/ الإسرائيلية/ حماس/ تل/ أبيب/ إسرائيل	التي
صواريخ	اطلقت/ إطلاق/ تل/ أبيب/ اتجاه	المقاومة/ عسقلان/ القبة/ الحديدية/ خمسة/ ثلاثة/ قصفت/ أربعة/ القسام/ كتائب/ القدس/ مدينة/ صواريخ/ إسرائيلية/ قالت	ــــــــــ
العسكرية	العملية/ العمليات/ الحملة	ــــــــــ	الإسرائيلية/ ضد/ منذ



الكلمة	المصاحبات اللفظية		
	تشترك بها القناتان	تكثر في الجزيرة	تكثر في العربية
المدنيين	استهداف/ حماية	الفلسطينيين/ قتل/ الأطفال/ حق	الضحايا/ الأبرياء/ معظم
الشعب	الفلسطيني/ العدوان	مطالب/ حقوق/ ضد	المصري
مدينة	جنوب/ حي/ غزة/ شرق/ في	رفح/ شرقي/ قصف/ خان/ يونس/ شمال/ وسط/ القدس/ الضفة/ تل/ أبيب/ صواريخ	الشجاعية/ استهدفت/ غارة

الجدول (٧): المصاحبات اللفظية للكلمات الأكثر تكرارا المشتركة في القناتين.

وكما يظهر في الجدول السابق، فإن القناتين لا تشتركان فقط في العديد من الكلمات الأكثر تكرارا، بل تشتركان أيضا في كثير من الكلمات المصاحبة لها. فـ(القصف) (الإسرائيلي) لـ(الفلسطينيين) من المحاور التي تبرز بشكل جلي في القناتين معا، مع التركيز على أهداف القصف وما ينتج عنه من ضحايا في الأرواح. ومع ذلك تختلف القناتان في الوصف المستخدم للتعبير عن ضحايا القصف. فالجزيرة تستخدم (استشهاد) ومشتقاتها، والعربية تستخدم (قتل) ومشتقاتها. وضمن المحاور المرتبطة بالقصف، تشترك القناتان في التركيز على الموضوع المتعلق بـ(استهداف) المدنيين و(حمايتهم)، كما يظهر من المصاحبات اللفظية لكلمة (المدنيين) التي تشير، كما ذكرنا سابقا، إلى إحدى الفئات الاجتماعية الرئيسية في الحدث. ومع أن القناتين تشتركان في استخدام كلمة (الحرب) بشكل متساو تقريبا، إلا أن الجزيرة تستخدم كلمة (العدوان) بشكل أكبر بكثير، في حين أن كلمة (الهجوم) أكثر بروزا في العربية. ويوضح الجدول (٨) الوزن النسبي لهذه الكلمات في القناتين.



الكلمة	التكرار في الجزيرة	التكرار في العربية	التكرار في القناتين	التكرار النسبي في الجزيرة	التكرار النسبي في العربية	التكرار النسبي في القناتين
الحرب	٧٨٥	١٦٦	٩٥١	٠,١٤٣٣١١	٠,١٠٤٨٩٦	٠,١٢٤١٠٤
العدوان	١٩٩٥	١٢٥	١٧٢٠	٠،٢٩١١٨٧	٠،٠٧٨٩٨٨	٠،١٨٥٠٨٨
الهجوم	١٥١	١٦١	٣١٢	٠،٠٢٧٥٦٧	٠،١٠١٧٣٧	٠،٠٦٤٦٥٢

الجدول (٨) التكرار النسبي لكلمتي (العدوان) و(الهجوم) في القناتين.

ومع أن هذه الكلمات هي الأكثر استخداما في القناتين لتسمية الحدث، إلا أنهما يستخدمان أيضا مصطلحات أخرى: (العملية/ العمليات/ الحملة) التي تصاحب كلمة (العسكرية)، كما رأينا في الجدول (٧). وكما يتضح في الجدول (٩)، فإن استخدام هذه المصطلحات قليل نسبيا ومتساو في القناتين معا. وبالرجوع إلى الكشاف السياقي تبين أن أغلب المواضع التي تستخدم فيها هذه المصطلحات تحكي وجهة النظر الإسرائيلية لرؤية الحدث.

المصطلح	التكرار في الجزيرة	التكرار في العربية	التكرار في القناتين	التكرار النسبي في الجزيرة	التكرار النسبي في العربية	التكرار النسبي في القناتين
العملية العسكرية	٧٦	٤٥	١٢١	٠،٠١٥٣٣٨	٠،٠٣٢٠٧١	٠،٠٢٣٧٠٥
العمليات العسكرية	٦٦	٣٢	٩٨	٠،٠١٣٣٢٠	٠،٠٢٢٨٠٦	٠،٠١٨٠٦٣
الحملة العسكرية	٣٢	١٣	٤٥	٠،٠٠٦٤٥٨	٠،٠٠٩٢٦٥	٠،٠٠٧٨٦١

الجدول (٩) استخدام مصاحبات (العسكرية) في القناتين.



وكما تشير المصاحبات اللفظية لكلمتي (المصرية) و(القاهرة)، فإن كلتا القناتين تشتركان في التركيز على (المبادرة المصرية) وما يصاحبها من (مفاوضات) من أجل التوصل إلى (وقف إطلاق النار). وقد قمنا باستخراج سلاسل الألفاظ الثلاثية الأكثر تكرارا في المدونة عموما، فحصلنا على النتيجة التالية في الجدول (١٠).

السلسلة اللفظية	التكرار في الجزيرة	التكرار في العربية	التكرار في القناتين	التكرار النسبي في الجزيرة	التكرار النسبي في العربية	التكرار النسبي في القناتين
قطاع غزة و	٧٦١	٢٧٦	١٠٣٧	٠،١٥٣٥٨٩	٠،١٩٦٧٠٧	٠،١٧٥١٤٨
على قطاع غزة	٥٥١	١٦٩	٧٢٠	٠،١١١٢٠٦	٠،١٢٠٤٤٨	٠،١١٥٨٢٧
في قطاع غزة	٥١٩	٢٦٩	٧٨٨	٠،١٠٤٧٤٧	٠,١٩١٧١٨	٠،١٤٨٢٣٣
وقف إطلاق النار	٤٩٤	٢١٨	٧١٢	٠،٠٩٩٧٠٢	٠،١٥٥٣٧	٠،١٢٧٥٣٦
المقاومة الاسلامية حماس	٤٧٧	٤٠	٥١٧	٠،٠٩٦٢٧١	٠،٠٢٨٥٠٨	٠،٠٦٢٣٨٩
حركة المقاومة الاسلامية	٤٦٥	٤٤	٥٠٩	٠،٠٩٣٨٤٩	٠،٠٣١٣٥٩	٠،٠٦٢٦٠٤

الجدول (١٠): السلاسل الثلاثية من الألفاظ الأكثر تكرارا في المدونة.

وكما يظهر من الجدول (١٠)، فإن (وقف إطلاق النار) من السلاسل ذات التكرار المرتفع في القناتين معا. ومع ذلك، فإن الوزن النسبي لسلسلة (وقف إطلاق النار) في العربية أكثر ارتفاعا منه في الجزيرة. وهذا يتفق مع ما ذكرناه سابقا من ورود كلمة (المبادرة) ضمن الكلمات الأكثر تكرارا في العربية. فالاهتمام بهذا المحور في العربية أكثر بروزا منه في الجزيرة. ومن ناحية أخرى، يوجد اختلاف واضح بين القناتين فيما تركزان



عليه من الموضوعات المرتبطة بـ(المبادرة). فمن النتائج اللافتة في الجدول (٧)، وجود كلمتين متضادتين ضمن المصاحبات اللفظية لكلمة (المصرية) في القناتين. فالعربية تنفرد بوجود تصاحب بين كلمتي (المصرية) و(دعم)، وأما الجزيرة، في المقابل، فتنفرد بوجود تصاحب بين كلمتي (المصرية) و(رفض). وبالرجوع إلى الكشاف السياقي، وجدنا أن كلمة (دعم) في العربية ترد في سياق الإعلان عن تأييد (المبادرة المصرية) ومساندتها من قبل عدد من الأطراف المعنية بالصراع محليا ودوليا. وأما كلمة (رفض) في الجزيرة فترد في سياق التعبير عن موقف (فصائل المقاومة الفلسطينية) من المبادرة. ويعرض الجدولان (١١) و(١٢) عينة من الكشاف السياقي لتصاحب هذه الكلمات في القناتين.

الجملة السابقة	الكلمة	الجملة اللاحقة
دعم عربي ل المبادرة	المصرية	لوقف اطلاق النار بغزة اعلن المتحدث ب اسم وزارة
الاميركي، ان وزير الخارجية كيري يعمل على دعم المبادرة	المصرية	في غزة.
الاوسط، عن ترحيب هـ و دعم هـ ل المبادرة	المصرية	ل التهدئة في غزة، و التي تهدف ل وقف
الخمسة عشر، اكد المجلس دعم هـ الكامل ل المبادرة	المصرية	و دعا الطرفين الى استئناف المفاوضات ل التوصل سريعا الى
الهاتفي على تاكيد تاييد السعودية و دعم ها ل المبادرة	المصرية	الرامية ل وقف اطلاق النار بين الجانبين الفلسطيني و الاسرائيلي
حل لغزة "، معبرا عن دعم باريس لـ الوساطة	المصرية	. و اكد هولاند ان " محمود عباس و السلطة



الجملة السابقة	الكلمة	الجملة اللاحقة
دولار. مؤكدا في كلمت هـ على ضرورة دعم الجهود	المصرية	ل وقف العدوان على غزة. و اوضح " نجتمع
عشر، اكد المجلس " دعم هـ الكامل ل المبادرة	المصرية	و دعا الطرفين الى استئناف المفاوضات ل التوصل سريعا الى
و كشف السيسي ان ايطاليا ابدت دعم ها ل المبادرة	المصرية	ل حل الازمة الراهنة في غزة، موضحا ان هذه
المتحدث ب اسم وزارة الخارجية المصرية ان الجهود و الاتصالات	المصرية	اثمرت عن دعم مجلس الجامعة العربية ل المبادرة المصرية ل

الجدول (١١) عينة من الكشاف السياقي للتصاحب بين كلمتي (دعم) و(المصرية) في العربية.

الجملة السابقة	الكلمة	الجملة اللاحقة
اخرين، و ذلك بعد رفض فصائل المقاومة مبادرة التهدئة	المصرية	و قال عضو اللجنة المركزية ل حركة التحرير الوطني
الاسرائيلي علي غزة، و اعلنوا رفض هم ل المبادرة	المصرية	ب اعتبار ها تصب في مصلحة اسرائيل و تساوي بين
الجناح العسكري ل حماس امس، رفض ها ل المبادرة	المصرية	، معتبرة ان ها " مبادرة ركوع و خنوع و
الشعبية ل تحرير فلسطين قد اعلنت ايضا رفض ها المبادرة	المصرية	، مؤكدة ان ها لا تلبي مطالب الشعب الفلسطيني،
المقاومة الفلسطينية. و اكد ان رفض حماس ل المبادرة	المصرية	ل التهدئة كان ل اعتبارات شكلية وجوهرية، " ف



الجملة السابقة	الكلمة	الجملة اللاحقة
الى استمرار رفض حركة حماس المبادرة. و تنص المبادرة	المصرية	على وقف ما سمي " الاعمال " العدائية "
ان ها س تدفع الثمن غاليا ل رفض ها المبادرة	المصرية	ل وقف اطلاق النار. و قال نتنياهو – خلال
على غزة ب سبب ما سمياه رفض حركة حماس المبادرة	المصرية	و استئناف ها اطلاق الصواريخ على المدن و المواقع الاسرائيلية
ل حماس. و ترى الصحيفة ان رفض حماس المبادرة	المصرية	ياتي من شعور ها ب ان ايادي كثيرة تخنق ها
خلال هذه الفترة، و طالب ب التخلي عن المبادرة	المصرية	بعد رفض ها من قبل كافة الفصائل الفلسطينية و الانتقال

الجدول (١٢) عينة من الكشاف السياقي للتصاحب بين كلمتي (رفض) و(المصرية) في الجزيرة.

وكما هو واضح، فإن القناتين تركزان على رؤيتين مختلفتين إزاء المبادرة. ففي الجزيرة تركيز على إبراز الرؤية الخاصة بـ(فصائل المقاومة الفلسطينية) وموقفها. ومما يصب في هذا الاتجاه أيضا ظهور كلمتي (مطالب) و(حقوق) ضمن المصاحبات اللفظية لكلمة (الشعب) في قناة الجزيرة (في إشارة إلى الشعب الفلسطيني). وتجدر الإشارة في هذا الاتجاه أيضا إلى مصاحبات كلمتي (صواريخ) و(الصواريخ)، كما يعرضها الجدول (٧). ففي الجزيرة عدد أكبر من المصاحبات التي ترد في سياق إبراز المجهود الحربي لفصائل المقاومة، وتركيز على التفاصيل في أنواع الصواريخ المستخدمة ومداها وأثرها، كما توضح عينة من الكشاف السياقي لكلمة (صواريخ) في الجزيرة في الجدول (١٣).

الجملة السابقة	الكلمة	الجملة اللاحقة
(حماس) – ان ها اطلقت ليل الخميس ثلاثة	صواريخ	بعيدة المدى على تل ابيب. ك ما قالت اسرائيل



الجملة السابقة	الكلمة	الجملة اللاحقة
حسب روايته. و كانت كتائب القسام قد اطلقت عدة	صواريخ	طويلة المدى من نوع " ام ٧٥ " في اتجاه
مطار تل ابيب، و اعترف الجيش الاسرائيلي ب سقوط	صواريخ	من غزة قرب مفاعل ديمونة النووي. و اضافت الكتائب
الكتائب انها قصفت ك ذلك مدينة تل ابيب ب اربعة	صواريخ	من طراز " ام ٧٥ "، و ب دورها
الجناح العسكري ل حركة الجهاد الاسلامي - انها اطلقت امس ثلاثة	صواريخ	من طراز " براق ٧٠ " على تل ابيب.
زيكيم. قرب ديمونة و قال الجيش الاسرائيلي ان ثلاثة	صواريخ	اطلقت من قطاع غزة تجاه بلدة ديمونة الاسرائيلية و ومفاعلها
صاروخا شمال مدينة القدس. و اقرت ب سقوط عشرة	صواريخ	ب اتجاه النقب اصاب احداها باحة منزل في المجلس الاقليمي اشكول
اسرائيل اجراءاتها الاحترازية في المطارات بعد اعتراض القبة الحديدية خمسة	صواريخ	كانت تستهدف مطار تل ابيب. و كان متحدث عسكري
الرحلات و تغيير المسار الجوي ل الطائرات المدنية بعد وصول	صواريخ	المقاومة الفلسطينية قرب المطار. و اثار سقوط الصواريخ على تل
تكاليف الصواريخ الاعتراضية او ما تسمى ب القبة الحديدية التي تعترض	صواريخ	المقاومة و التي تكلف الاقتصاد الاسرائيلي خسائر كبيرة لا يقوى على

الجدول (١٣) عينة من الكشاف السياقي لكلمة (صواريخ) في الجزيرة.

وفي سياق المصاحبات اللفظية للكلمات الأكثر تكرارا أيضا، نود الإشارة إلى ملاحظة



ختامية تتشمل في انفراد الجزيرة بوجود تصاحب وثيق بين (مصر) و(معبر رفح)، وهو ما يشير إلى بروز موضوع المعبر في التغطية الصحفية للجزيرة، بخلاف العربية.

ورغم أهمية الكلمات الأكثر تكرارا في الدلالة على بعض المؤشرات العامة، إلا أنها وحدها ليست كافية في توضيح ما يميز الاستعمال اللغوي في كل قناة. ولذا قمنا باستخراج الكلمات المميزة في المدونتين تبعا لعدد من المقاييس والمحددات التي شرحناها في الجزء السابق. ويعرض الجدولان (١٤) و(١٥) الكلمات المميزة في مدونتي الجزيرة والعربية على التوالي.

الجزيرة/ المقاومة/ الاحتلال/ استشهاد/ الصحيفة/ استشهد/ شرقي/ الكاتب/ المسحال/ جنوبي/ تموز/ شمالي/ تامر/ يرى/ القناة/ الصحف/ الدحدوح/ الانقلاب/ المراسل/ مائة/ المظاهرة/ وائل/ الشهيد/ الحي/ يديعوت/ المتظاهرون/ الصحفيين/ شعارات/ مقالا/ اونروا/ مقال/ استشهدوا/ الشعبية/ النشر/ المعلق/ البلدة/ كرام/ مظاهرة/ شهيد/ تساءل/ صحف/ تايمز/ الصمت/ استنزاف/ نابلس/ بلدات/ التبرع/ المالكي/ يضيف/ سما/ خمسين/ تضيف/ سلطات/ الكاتبة/ النقب/ محفوظة/ مائتي/ افتتاحية/ اهم/ احرونوت/ الياس/ المقاومين/ تشرين/ الهاون/ الاسير/ مرسي/ أبناءئ/ ديفد/ التحالف/ اندبندنت/ الرجوب/ خمسمائة/ الحرم/ الناصر/ يعكس/ تحرير/ ذي/ العطية/ مؤيدة/ العدالة/ خيار/ غارديان/ القطري/ الموحد/ عصر/ المكرمة/ بحي/ انحياز/ هجومية/ العسكر/ البديري/ العزل/ بينت/ عكيلة/ عاموس/ جيفارا/ صحفية/ كلما/ العمق/ الموافق/ الميدان

الجدول (١٤) الكلمات المميزة في الجزيرة.

فرانس/ القتلى/ برس/ المئة/ الحرمين/ عبدالعزيز/ خادم/ قتيلا/ ليبيا/ صحافي الصحافيين/ عبدالله/ الشريفين/ النزاع/ الإيراني/ الملك/ شكري/ الصحافي/ عبدالفتاح/ الجزائرية دلو/ الاوقية/ كلينتون/ برنت/ جنبلاط/ الليبي/ البرميل/ طبي/ داعش/ الذهب الليبية/ الحرس/ المسلسلات/ الفيديو/ بوتفليقة/ المصغرة/ ديفيد/ الجوار/ سعر/ امدادات ولي/ زايد/ غالبية/ الازهر/ الأمير/ البلدين/ القمة/ تظاهرة/ النفط/ العاهل روحاني/ الدائم/ كوريا/ الاسد/ العهد/ انفجارات/ المسلح/ توفي/ المندوب/ التراب قوافل/ الحوار/ تعقد/ المجموعات/ مساعد/ الجليل/ الغاشم

الجدول (١٥) الكلمات المميزة في قناة العربية.



ومع أهمية المقاييس الكمية في استخراج الكلمات المميزة، إلا أنها وحدها ليست كافية. ولذا فقد قمنا بمراجعة الكلمات المميزة في كل مدونة كلمة كلمة لمعرفة عدد مرات ورودها، والسياقات التي وردت فيها عبر استخدام الكشاف السياقي. فالكلمات التي وردت مئة مرة فأكثر رجعنا إلى عينة عشوائية من كشافها السياقي، وأما التي وردت أقل من ذلك فقد رجعنا إلى كامل كشافها السياقي من أجل تحديد معانيها ودواعي استعمالها. وقد استطعنا بذلك تحديد الكلمات المرتبطة بالحدث لأننا وجدنا العديد من الكلمات التي وردت استطرادا في سياق التعليق على أحداث أخرى غير متصلة بموضوع غزة، ويكثر ذلك في أخبار قناة العربية. فعلى سبيل المثال، وردت كلمة (ليبيا) ضمن الكلمات المميزة لقناة العربية. وبالرجوع إلى الكشاف السياقي لهذه الكلمة وجدنا أن كل المواضع التي وردت فيها هذه الكلمة ليس لها علاقة بالأحداث المتصلة بموضوع غزة. ويعرض الجدول (١٦) عينة من الكشاف السياقي لكلمة ليبيا في قناة العربية.

الجملة السابقة	الكلمة	الجملة اللاحقة
كما تم تناول الاوضاع في كل من سوريا و	ليبيا	وانعكاساتها على كل من مصر والمملكة العربية
اضافة الى دعوة " كافة الاطراف السياسية في	ليبيا	الى حل خلافاتها عبر الحوار وانتهاج مسار توافقي
التي تواجه ها دول الجوار جراء تردي الاوضاع الامنية في	ليبيا	، والتاكيد على ضرورة معالجة بؤر الارهاب فيها
التعاون مع مصر ل مواجهة الارهاب.. الوضع في	ليبيا	يحتاج ل تدخل سريع وقوي في هذا البلد "
س يبحث مع هـ تطورات الوضع في غزة وفي	ليبيا	ورجحت مصادر مسؤولة ان تكون الجزائر ب صدد
الامنية. واشار جبريل الى ان ما يجري في	ليبيا	من تطورات متصارعة ل هـ تداعيات سلبية على مصر،

الجملة السابقة	الكلمة	الجملة اللاحقة
طلب المساعدة الدولية منعا ل تطور الازمة، قائلا:	ليبيا	حصلت على استقلال ها ب قرار من الجمعية العامة ل
غزة، و ان كانت الصراعات المحتدمة في العراق و	ليبيا	وهما من اعضاء اوبك لم تؤثر بعد على انتاج
ب وزارة النفط الليبية، امس الثلاثاء، ان انتاج	ليبيا	من النفط لايزال عند نحو ٥٠٠ الف برميل يوميا

الجدول (١٦): عينة من الكشاف السياقي لكلمة (ليبيا) في قناة العربية.

وبناء على استعراض الكشاف السياقي لجميع الكلمات المميزة في القناتين، قمنا باستبعاد الكلمات التي ليس لها علاقة بأحداث غزة، واكتفينا بالكلمات المتصلة بالموضوع، مع تصنيفها إلى أربعة محاور رئيسية تبعا لدلالاتها، كما فعلنا في الكلمات الأكثر تكرارا. وبخلاف الكلمات الأكثر تكرارا، فإن الاختلاف بين القناتين يظهر بشكل أوضح في الكلمات المميزة. ويحتوي الجدول (١٧) على ملخص بالكلمات المميزة في القناتين مصنفة ضمن المحاور التي تنتمي إليها.

المحور الذي تنتمي إليه الكلمات		الكلمات المميزة	
		مميزة في الجزيرة	مميزة في العربية
أطراف الصراع/ وأماكنه/ والمعنيون به	أقاليم	———	———
	دول (أو كلمات منسوبة لها)	سلطات/ القطري	———
	مدن	نابلس/ النقب/ الحرم/ بحي	الجليل

المحور الذي تنتمي إليه الكلمات		الكلمات المميزة	
		مميزة في الجزيرة	مميزة في العربية
أطراف الصراع/ وأماكنه/ والمعنيون به	منظمات	اونروا	الأزهر/ مندوب
	قطاعات وفئات اجتماعية مرتبطة بالحدث	الأسير/ العزل	————
	كلمات تشير إلى مواضع غير محددة	شرقي/ جنوبي/ شمالي/ البلدة/ بلدات	————
كلمات تشير إلى وصف مباشر للحدث		المقاومة/ الاحتلال/ استشهاد/ استشهد/ الشهيد/ استشهدوا/ شهيد/ استنزاف/ المقاومين/ الهاون/ تحرير/ هجومية/ العمق	قتيلا/ القتلى/ النزاع/ انفجارات/ المسلح/ توفي/ الغاشم/ طبي/ الجوار
كلمات تشير إلى مواقف مصاحبة للحدث ومرتبطة به		الانقلاب/ مظاهرة/ المظاهرة/ المتظاهرون/ شعارات/ مؤيدة/ التبرع/ الشعبية/ الصمت/ انحياز/ العدالة	القمة/ تظاهرة/ شكري



المحور الذي تنتمي إليه الكلمات	الكلمات المميزة	
	مميزة في الجزيرة	مميزة في العربية
كلمات تشير إلى التغطية الصحفية للحدث	الجزيرة/ الصحيفة/ الكاتب/ تامر/ المسحال/ القناة/ الصحف/ الدحدوح/ المراسل/ وائل/ يدعوت/ الصحفيين/ مقال/ تساءل/ المعلق/ كرام/ صحف/ تايمز/ يضيف/ تضيف/ الكاتبة/ افتتاحية/ احرونوت/ إلياس/ إندبندنت/ غاردين/ البديري/ جيفارا/ عكيلة/ بينت/ صحفية/ الموافق/ مرسي/ العطية/ العدالة	فرانس/ برس/ صحافي/ الصحافي/ الصحافيين

الجدول (١٧) المحاور التي تنتمي لها الكلمات المميزة في القناتين.

وكما هو واضح في الجدول السابق، فإن الكلمات المميزة في (الجزيرة) أكثر بكثير منها في (العربية). ويعود هذا التفاوت إلى سببين رئيسيين. أحدهما إحصائي يرجع إلى تفاوت العدد الكلي للكلمات في القناتين بالنسبة إلى إجمالي المدونة. فكلمات الجزيرة، كما ذكرنا في وصف البيانات سابقا، تمثل ما يقارب ٦٩٪ من العدد الكلي لكلمات المدونة بمجموعها. والسبب الثاني يرجع إلى نوعية المحتوى الذي تقدمه القناتين. فالتغطية الصحفية في قناة الجزيرة، كما مر معنا في النتائج التي استعرضناها حتى الآن، تحتوي على قدر أكبر من التفصيل مقارنة بقناة العربية. ومن مظاهر التفصيل في الكلمات المميزة لقناة الجزيرة، على سبيل المثال، ورود عدد أكبر من الكلمات التي تشير إلى مواضع متعددة ومختلفة، بالإضافة إلى ظهور عدد أكبر من أسماء المراسلين في الميدان وأسماء



الصحف التي تعرض وجهات نظر مختلفة حول الحدث.

وبطبيعة الحال فإنه لا يمكننا التوقف عند كافة الكلمات المميزة في القناتين، ولذا فإننا سنقتصر على أبرز ما يمكن ملاحظته في المحور الذي يحتوي على (الكلمات التي تشير إلى وصف مباشر للحدث) والمحور الذي يحتوي على (الكلمات التي تشير إلى مواقف مصاحبة للحدث ومرتبطة به) في الجدول (١٧). أما بالنسبة إلى الكلمات التي تصف الحدث بشكل مباشر، فيمكن الإشارة إلى ثلاث ملاحظات أساسية. أولها وجود اختلاف واضح في الكلمات المستخدمة لوصف ضحايا الحرب من الفلسطينيين. فالجزيرة تستخدم (شهيد) ومشتقاتها، والعربية تستخدم (قتيل) ومشتقاتها. وهذا يؤكد ما ذكرناه سابقا في التعليق على الجدول (٧)، فهذان الوصفان المختلفان لا يظهران ضمن المصاحبات اللفظية للكلمات الأكثر تكرارا فقط، بل هما من الكلمات المميزة لكل من القناتين مما يدل على تباين واضح بين القناتين من هذه الناحية. وأما الملاحظة الثانية فهي استخدام الجزيرة لكلمات تحمل تقييما معينًا في توصيف الصراع، مثل: (المقاومة، الاحتلال، تحرير)، في مقابل استخدام كلمة (النزاع) في قناة العربية. وأما الملاحظة الثالثة فتتمثل في ورود عدد من الكلمات، مثل: (استنزاف، الهاون، هجومية، العمق) التي تدخل ضمن التفصيل في إبراز المجهود الحربي لفصائل المقاومة، في مقابل اكتفاء العربية باستخدام كلمات عامة مثل (انفجارات). وهذه الملاحظة الأخيرة تؤكد ما ذكرناه سابقا عن اهتمام الجزيرة بهذا الجانب. فكلمة (استنزاف)، على سبيل المثال، وردت في الجزيرة في (٥٠) موضعا تشير في معظمها إلى (تخوف) إسرائيل من الحرب، واستعداد فصائل المقاومة لها. وأما في العربية، فلم ترد هذه الكلمة مطلقا. ويحتوي الجدول (١٨) على عينة للكشاف السياقي لكلمة (استنزاف) في قناة الجزيرة.

الجملة السابقة	الكلمة	الجملة اللاحقة
و عسكريون اسرائيليون ان تل ابيب تتحسب من دخول حرب	استنزاف	مع المقاومة الفلسطينية ب قطاع غزة، نظرا ل الخسائر
الاستخبارات يوفال شطاينتس، عن مخاوف اسرائيل من " حرب	استنزاف	" في حال عدم احراز تسوية سياسية في الوقت الراهن



الجملة السابقة	الكلمة	الجملة اللاحقة
معلقين اخرين يقولون ان اسرائيل " تخشى الدخول في حرب	استنزاف	تنهك ها وسط انشغال العالم ب ما يجري في العراق
العاشرة " هل س تقبل اسرائيل ب الدخول في حرب	استنزاف	جديدة قد تمتد ١٤ عاما ؟ ". و يشير
ان المقاومة الفلسطينية " قادرة اليوم على الاستمرار ب حرب	استنزاف	حتى لو انخفض معدل صواريخ ها اليومي من ١٥٠ الى
" تعي " الاثمان الاقتصادية و السياسية المترتبة على حرب	استنزاف	، مشيرا الى ان جنوب اسرائيل " معطل و سكان
فارس ل الجزيرة نت ان الخسائر الاقتصادية الكبيرة ل حرب	استنزاف	ك هذه " تحول دون عودة عشرات الاف العمال ل
حماس) تل ابيب من ان ها س تواجه حرب	استنزاف	في حال رفضت تلك المطالب. و قال رئيس الوفد
، مشيرين الى ان تل ابيب لا تتحمل تبعات حرب	استنزاف	في قطاع غزة. و اضاف المسؤولون ان تنتياهو قد
الاسلامية (حماس) نجحت في جر اسرائيل ل حرب	استنزاف	مستمرة منذ ٥٠ يوما، دون المساس بقدراتها

الجدول (١٨): عينة من الكشاف السياقي لكلمة (استنزاف) في الجزيرة.

وأما من ناحية الكلمات التي تشير إلى مواقف مصاحبة للحدث، فيمكن الإشارة إلى ثلاث ملاحظات أيضا. أولها وجود عدد من الكلمات في الجزيرة تركز بشكل أكبر على ردود الأفعال (الشعبية) تجاه الحدث، مثل: (مظاهرة، المظاهرة، المتظاهرون، شعارات، مؤيدة، التبرع، الشعبية)، في مقابل تركيز العربية بشكل أكبر على الموقف الرسمي الذي تعبر عنه كلمتا: (القمة) و(شكري) (التي تشير إلى وزير الخارجية المصري سامح شكري). وهذه الملاحظة تؤيد ما ذكرناه سابقا من اهتمام العربية بالمواقف الرسمية (الداعمة) للمبادرة العربية إقليميا ودوليا. وأما الملاحظة الثانية فهي



بروز كلمة (انقلاب) في قناة الجزيرة ضمن الكلمات التي تشير إلى المواقف المصاحبة للحدث. فالكلمة ترد في سياق وصف الوضع السياسي في مصر، وعلاقته بالأحداث في غزة. وبروز هذه الكلمة في الجزيرة يرتبط بما ذكرناه سابقا حول وجود تصاحب لفظي واضح بين (مصر) و(معبر رفح) في قناة الجزيرة. وأما الملاحظة الثالثة فهي ورود عدد من الكلمات في قناة الجزيرة تشير إلى الموقف العام في المجتمع الدولي إزاء الحدث الفلسطيني، مثل: (الصمت، انحياز، العدالة). وهذا المحور شبه غائب تقريبا في قناة العربية. فكلمة (الصمت)، على سبيل المثال، وردت في الجزيرة في (٦٩) موضعا في سياق وصف الموقف العام من الحدث الفلسطيني، ومصاحباتها تتضمن الكلمات التالية: (الدولي، العربي، إزاء، تجاه، المجازر). وأما في العربية، فلم ترد هذه الكلمة إلا في ثلاثة مواضع فقط في سياق كلمة نصية لـ (الملك عبدالله بن عبدالعزيز-رحمه الله-) في التعليق على الحدث. ويحتوي الجدول (١٩) على عينة عشوائية من الكشاف السياقي لكلمة (الصمت) في قناة الجزيرة.

الجملة السابقة	الكلمة	الجملة اللاحقة
ملك السعودية يندد ب	الصمت	الدولي تجاه مجازر غزة قال الملك السعودي عبد الله بن
الوقفة ب حرق علم كبير ل اسرائيل، منددين ب	الصمت	العربي، ك ما رددوا هتافات تؤكد ان المقاومة هي
ضد المدنيين الابرياء في غزة، و ندد ب	الصمت	الدولي تجاه المجازر في فلسطين. واضاف
مثل غزة. يشار الى ان منظمة كسر	الصمت	انشئت عام ٢٠٠٤، و قد اطلق ها مجموعة
ب ادانة المجزرة التي ترتكب ها اسرائيل، مؤكدة ان	الصمت	الدولي هو الذي يشجع هذه الحكومة على المضي في جريمة
هـ الصمت المخجل عما يجري في غزة، معتبرين هذا	الصمت	يرقى الى مستوى التواطؤ مع الاسرائيليين، و طالبوا ب



الجملة السابقة	الكلمة	الجملة اللاحقة
بعد ها مباشرة بدا ب ارتون انتقاد اسرائيل، و	الصمت	الدولي عن مجازر ها، واصفا الغرب ب محاباة اسرائيل
، و رفع المشاركون لافتة ضخمة كتب علي ها "	الصمت	يقتل "، في اشارة الى استنكار هم ل الصمت
الصمت يقتل في اشارة الى استنكارهم ل	الصمت	الاوروبي تجاه ما يحدث في غزة. وشهدت العاصمة
الاسرائيلية ب الهمجية والمتوحشة، ك ما نددوا ب	الصمت	العربي والدولي على الجرائم الاسرائيلية. من جهة ها

الجدول (١٩) عينة عشوائية من الكشاف السياقي لكلمة (الصمت) في قناة الجزيرة.

هذه هي أبرز النتائج التي حصلنا عليها من خلال تحليل المدونة اللغوية للتغطية الإخبارية في قناتي الجزيرة والعربية لأحداث غزة في الفترة من ٧/ ٧/ ٢٠١٤م إلى ٢٧ / ٨ / ٢٠١٤م. وفي القسم التالي سنقوم بمناقشة هذه النتائج بشكل عام في ضوء الإطار النظري العام للتحليل النقدي للخطاب، مع محاولة ربط هذه النتائج بنتائج الدراسات السابقة التي أشرنا لها سابقا.

٦. ٧. مناقشة النتائج

لقد اقتصرنا في هذه الدراسة على استعراض الكلمات الأكثر تكرارا والكلمات المميزة ومصاحباتها في القناتين. ومن خلال المقارنة بين الكلمات المستخدمة في القناتين، استطعنا الحصول على صورة وافية -كما نعتقد- للمحاور العامة التي تمثل نوعية الخطاب المستخدم في كل منهما لتناول أحداث غزة. وكما أشرنا في القسم السابق؛ ثمت نقاط تشابه بين القناتين في بعض الجوانب، بالإضافة إلى وجود العديد من الاختلافات في جوانب أخرى. وسنحاول هنا تلخيص أبرز النتائج، بدءا بأوجه الشبه، ثم الاختلافات، وبعد ذلك نعلق على المنهجية العامة للدراسة الحالية ومخرجاتها بالنسبة إلى إمكانات البحث اللغوي في التحليل النقدي للخطاب.

أما بالنسبة إلى أوجه الشبه بين القناتين، فيوجد عدد كبير من الكلمات المشتركة



بينهما، سواء في الكلمات الأكثر تكرارا أو في مصاحباتها اللفظية. ومن الملاحظات المشتركة أن القناتين معا تركزان بشكل أساسي على تقديم الأخبار التي تحكي ما يقع من أحداث، وذلك مثل القصف وما ينتج عنه من خسائر وضحايا في الأرواح والممتلكات في الجانبين، بالإضافة إلى التحركات السياسية المختلفة للتعامل مع الموقف. فالكلمات التي استعرضناها في القسم السابق تعبر عن المحاور الأساسية في الصراع، وتشرح ما يشتمل عليه من أبعاد مختلفة بشكل متوازن إلى حد كبير. ومن أوجه الشبه البارزة بين القناتين تركيزهما على البعد الإنساني للحدث، ولذا تظهر فيهما الكثير من الكلمات المتصل بهذا المحور، مثل: (استهداف) (المدنيين) (العزّل) و(الضحايا) (الأبرياء)، بما فيهم (الأطفال) و(العوائل)، و(حمايتهم). ورغم اشتراكهما في التركيز على هذا المحور، إلا أنهما، كما قلنا سابقا، تركزان في المقام الأول على نقل الوقائع. فمن الملاحظات الأساسية في الكلمات التي استعرضناها في القسم السابق خلوها من الألفاظ العاطفية التي تحمل دلالات انفعالية لا يقتضيها النقل المجرد للخبر. فالقناتان من هذه الناحية بينهما تشابه كبير.

وهذه الملاحظات حول التشابه بين القناتين في الجوانب التي ذكرناها تتفق مع النتائج التي انتهت إليها بعض الدراسات السابقة (انظر مثلا: البغدادي، ٢٠١٠). وتتفق كذلك مع الملاحظة التي انتهى إليه فان ديك (van Dijk 1988)، في دراسته الرائدة عن خطاب الأخبار، وإشارته إلى عدم وجود اختلافات حادة في نقل وسائل الإعلام الكبرى للمحاور الأساسية في الأخبار الدولية. ويعزو فان ديك (van Dijk 2009: 198) هذا التشابه العام «إلى نجاح وكالات الأنباء العالمية التي تستند إليها معظم الأخبار في خلق صيغة عالمية موحدة في نقل الأحداث». وكما رأينا سابقا في استعراض الدراسات السابقة، فإن الاختلافات في وسائل الإعلام الكبرى أضيق بكثير منها في الوسائل الإعلامية الحزبية والمحلية التي تعكس وجهة نظر المؤسسات التي تنتمي إليها بشكل مباشر، وتتسم بنوع من التحيز الواضح في التعليق على الأحداث (انظر مثلا: بصيص ٢٠٠٧؛ أبو مزيد ٢٠١٣).

ورغم ما بين القناتين من تشابه مجمل في نقل الوقائع، فإن ذلك لا يعني أنهما ينطلقان من رؤية موحدة. فالكلمات التي استعرضناها في القسم السابق تعبر عن اختلافات واضحة في العديد من الجوانب. ويمكن تلخيص أوجه الاختلاف بين القناتين في



جانبين أساسيين: (١) الإطار العام لتقديم الحدث، و(٢) نوعية الكلمات التي تنتقيها كل قناة لتوصيف بعض الجوانب المحددة. أما من ناحية الإطار العام لتقديم الحدث، فالعربية تكتفي غالبا بإيراد المعلومات الرئيسية للوقائع، مع التركيز على إبراز الرؤية الرسمية التي تصدر عنها الدول والمنظمات الدولية الأخرى (مثل: الجامعة العربية ومنظمة الأمم المتحدة) في التعامل مع الحدث. وأما الجزيرة فتقدم الحدث من منظور يهتم بالوصف والسرد التفصيلي للوقائع وما تتضمنه من جوانب مختلفة، مع الحرص على وضع الحدث ضمن السياق الكلي الذي يحيط به محليا وعالميا. وهذه النتيجة تتفق إلى حد كبير مع النتيجة التي توصلت إليها ماهيناز محسن (محسن ٢٠٠٩) في دراستها المقارنة عن القناتين في تغطيتهما لأحداث غزة في عام ٢٠٠٨م.

ومن المؤشرات التي توضح اهتمام الجزيرة بالتفاصيل كثرة المفردات التي تشير إلى مواضع متعددة داخل فلسطين، بالإضافة إلى تميزها بورود عدد كبير من المفردات التي تشير إلى التغطية الميدانية من قبل المراسلين، وإلى مصادر صحفية متنوعة تنظر إلى الحدث من زوايا مختلفة، بخلاف العربية التي تقل فيها المفردات المرتبطة بهذا الجانب. وهذا يخالف إلى حد ما ورد في دراسة ريما البغدادي (٢٠١٠) حول اعتماد القناتين في تغطيتهما لأحداث غزة عام ٢٠٠٨م على مصادرهما الذاتية في نقل الخبر. فلم تظهر في الكلمات المميزة للعربية مفردات تشير إلى التغطية الميدانية من قبل المراسلين، بخلاف قناة الجزيرة التي ظهر فيها عدد كبير من الكلمات المميزة ضمن هذا المحور.

ومن الجوانب التي ترتبط بالإطار العام لتقديم الحدث وتوضح الاختلاف بين القناتين من هذه الناحية وجود عدد كبير من الكلمات التي تصف الفعاليات الشعبية المتصلة بالحدث في الجزيرة، مثل: (المظاهرات) و(المسيرات) و(حملات التبرع) و(الشعارات المؤيدة). وهذا يدل على اهتمام الجزيرة بهذا الجانب باعتباره أحد المظاهر الأساسية للحدث، بخلاف العربية التي لا تولي عناية كبيرة بهذا الجانب، وتركز بشكل أكبر على الجانب الرسمي، حيث تبرز فيها كلمات من قبيل: (المبادرة العربية)، و(القمة العربية) و(مجلس الأمن) و(الأمم المتحدة)، ونحوها. ومما يتصل بالإطار العام لتقديم الحدث أيضا؛ اهتمام الجزيرة بإبراز دور (فصائل المقاومة الفلسطينية) باعتبارها أحد الأطراف الفاعلة في الحدث، والتأكيد على موقفها في (رفض) (المبادرة)، وأثر عملياتها (الهجومية) في (العمق الإسرائيلي)، بخلاف العربية التي لا يظهر في كلماتها تركيز على



هذا الجانب، ويظهر بدلا من ذلك تركيزها على (دعم) (المبادرة) من قبل الأطراف العربية والدولية المعنية بالحدث في سياق اهتمامها بالموقف الرسمي من الحدث.

ومن مظاهر اختلاف القناتين في الإطار العام لتقديم الحدث أيضا عدم اقتصار الجزيرة على النقل المجرد للوقائع، بل تهتم أيضا بما يحيط بها من سياق وظروف وملابسات سياسية واجتماعية وتاريخية، قد تكون خفية في بعض الأحيان. ولذا تظهر فيها كلمات من قبيل: (العدالة) و(مطالب) (الشعب الفلسطيني) و(حقوقه). ومما يصب في هذا الاتجاه أيضا ظهور كلمات تعبر بشكل سلبي عن موقف المجتمع الدولي، مثل: (الانحياز)، و(الصمت). ومع أن القناتين تبديان اهتماما كبيرا بـ(مصر) ودورها في التعامل مع الحدث، إلا أنهما يختلفان جدا في الجوانب التي يركزان عليها في الدور المصري. وكما فصلنا في القسم السابق، فإن العربية تقتصر تقريبا على الدور الرسمي الذي يعمل على (التهدئة) و(دعم) (المبادرة العربية) لـ(وقف إطلاق النار). وأما الجزيرة، فتركز أكثر على بعض الفعاليات الشعبية كـ(المظاهرات) و(المسيرات) (المناهضة) لـ(الانقلاب). وربما ينطوي ذلك على إقحام لرؤية سياسية معينة تجاه الوضع السياسي في مصر في أحداث غزة.

وكما ذكرنا، لا يقتصر الاختلاف بين القناتين على اختلافهما في الإطار العام لتقديم الحدث، بل يشمل أيضا اختلافهما في نوعية الكلمات التي تنتقيها كل قناة لتوصيف بعض الجوانب. فيظهر في الجزيرة عدد من الكلمات المركزية والرئيسية التي تتضمن حمولات ثقافية وتاريخية، وتعبر ضمنا عن انتماء معين ورؤية إزاء الحدث. وأما في العربية فيظهر العدول عن هذه النوعية من الكلمات إلى كلمات أخرى أكثر عمومية، ولا تشي بانتماء معين. ومن هذا القبيل، على سبيل المثال، استخدام كلمة (شهيد) ومشتقاتها في الجزيرة، في مقابل استخدام كلمة (قتيل) ومشتقاتها في العربية. فكلمة (شهيد) نعت يستند إلى موروث ديني، ويتضمن صفة المدح لمن فقد حياته في قتال مشروع. فهذه الكلمة موحية وتعبر في وقت واحد عن الثناء على المقتول وإدانة القاتل، بخلاف كلمة (قتيل) التي لا تحمل أي دلالة أخرى سوى التعبير عن حدوث فعل القتل. ومن التقابلات الخفية بين الكلمتين أن (شهيد) من (فعيل) التي تتضمن معنى (فاعل) و(مفعول) في نفس الوقت، بخلاف كلمة (قتيل) التي لا تحتمل إلا معنى (مفعول). فـ(الشهيد) تحيل إلى (شاهد) و(مشهود) التي تتضمن الدور الفاعل للشهيد في التضحية بالنفس والشهادة



له بالمنزلة العالية. وأما (قتيل) بمعنى (مقتول) فلا تتضمن إلا التأثر السلبي بالفعل الذي يقع من الخارج. وهذا المعنى نفسه ينطبق أيضا على الفعلين اللذين وردا في القناتين: (اسْتَشْهَد) (في الجزيرة) و(قُتِل) (في العربية). فـ(اسْتَشْهَد) مبني للمعلوم على صيغة (استفعل) التي تدل على الطلب، وتتضمن معنى الإرادة ممن وقع عليه الفعل [1]، بخلاف (قُتِل) المبني للمجهول، ولا يتضمن إلا معنى التأثر السلبي.

وأما من ناحية الكلمات المستخدمة في الإشارة إلى الحدث ككل، تشترك القناتان في استخدام عبارة (الحرب على غزة). ويجمع بينهما استخدام حرف الجر (على) الذي يدل على الاستعلاء، ويشير إلى أن الحرب تشن من طرف على طرف آخر على نحو غير متكافئ. ومن الواضح أن هذه العبارة تتضمن معنى مقصودا يختلف عن المعنى في عبارة (الحرب في غزة) أو (حرب غزة) التي تستخدم في وسائل الإعلام الغربية (انظر مثلا: Kandil 2009). ومع تشابههما هنا، فإنهما تختلفان في الكلمة التي تستخدمانها لتوصيف السلوك الإسرائيلي. فالجزيرة، كما رأينا، تستخدم (العدوان)، بينما تستخدم العربية (الهجوم). و(العدوان) كلمة تدل بصيغتها وحروفها على المبالغة في التعدي، ولذا فإنها تتضمن تقييما سلبيا، بخلاف كلمة (الهجوم) التي تمثل مصطلحا عسكريا يعبر عن تحرك معين، ولا تحمل أي تقييم تجاه الحدث. وكلمة (العدوان) كذلك من الكلمات المتأصلة في الخطاب السياسي العربي منذ المراحل المبكرة للصراع العربي الإسرائيلي (انظر مثلا: عكاشة، ٢٠٠٥)، بخلاف كلمة (الهجوم) التي تمثل كلمة طارئة ومقترضة من الخطاب الدولي في وسائل الإعلام الغربية لترجمة كلمة (attack) (انظر مثلا: Kandil 2009). وبالمثل، فإن الجزيرة تستخدم كلمتي (المقاومة) و(الاحتلال). وكلاهما يدلان على تقييم معين، ومن الكلمات المتأصلة في الخطاب السياسي العربي، بخلاف كلمة (النزاع) التي تقابل كلمة (conflict) في وسائل الإعلام الأجنبية، وتستخدمها العربية على سبيل المجاراة لها في توصيف الصراع. وبطبيعة الحال، فإن الاختلاف في استخدام هذه الكلمات ليس اختلافا لغويا مجردا، بل يعكس موقفا أيديولوجيا معين في

١- لابد من الإشارة هنا إلى حدوث بعض التغير في الاستخدام المعاصر للفعل (استشهد) فأصبح يستخدم بصيغة المبني للمجهول. وهذا مخالف للقواعد المعيارية التي لا تجيز ذلك إلا في الفعل المتعدي، ولكنه أيضا لا يؤثر على التحليل الذي ذكرناه. فاستشهد المبنية للمجهول تدل على الطلب وتتضمن الإشارة إلى الإرادة أيضا.



رؤية العالم. ونظرا لضيق المساحة فإنه لا يمكننا استعراض المزيد من الأمثلة، ولذا فإننا سنكتفي بهذا القدر والذي نحسب أنه قدم تصورا وافيا.

٦. ٨. الخاتمة

إن نتائج الدراسة الحالية ومناقشتها تظهر بجلاء أهمية استثمار المدونة في دراسات تحليل الخطاب، والخطاب النقدي بوجه أخص. إذ لا يقتصر دور المدونة اللغوية على فتح آفاق جديدة في التحليل اللغوي فحسب؛ بل أسهمت من خلال الإمكانات الحاسوبية إلى البعد عن الذاتية والانتقائية ومحاولة الوصول إلى قدر أكبر من الموضوعية مقارنة بالدراسات التقليدية في التحليل النقدي للخطاب. كما أن استثمار المدونة اللغوية يمكن الباحث من الوصول إلى نتائج أكثر موضوعية وبيانات بحثية يستطيع من خلال استقرائها أن يبني تصوره الذاتي حيال موضوع الدراسة دون الاعتماد على رؤى بحثية مسبقة.

وفي الدراسة الحالية ومن خلال استثمار المدونة اللغوية المدروسة للقناتين تم الوصول إلى الكلمات الأكثر تكرارا، والكلمات المميزة، والكشف عن مصاحباتها اللفظية. وتلك البيانات ساعدت الباحثين في بناء حقول دلالية تنتمي لها تلك الكلمات بطريقة استقرائية لنتائج البحث، والتي أظهرت أنه يوجد أوجه للاتفاق، وأجه أخرى للاختلاف تنطلق من الإطار العام لتقديم الحدث، والتفضيلات اللغوية لتوصيف بعض جوانبه.

ويقترح الباحثان إجراء دراسات أخرى في السياق ذاته (الاحتلال الإسرائيلي لفلسطين) من خلال استثمار مدونة لغوية أكبر. ويمكن الإفادة من أنماط مختلفة من الخطاب الإعلامي (برامج تلفازية، نشرات إخبارية مقروءة، برامج حوارية، مواقع الإنترنت، مشاركات القنوات في وسائل التواصل الاجتماعي) ومقارنتها من خلال بعد زمني أكبر وذلك للوصول إلى نتائج أكثر عمقا وتمثيلا.

قائمة المراجع

أولا: المراجع العربية

- أبو مزيد؛ رجاء. (٢٠١٣). الخطاب الصحفي الفلسطيني نحو قضية المصالحة الفلسطينية: دراسة وصفية. رسالة ماجستير غير منشورة. الجامعة الإسلامية بغزة.
- براون، ج. يول، ج. (١٩٩٧). تحليل الخطاب. ترجمة: محمد لطفي الزليطني ومنير التريكي. الرياض: جامعة الملك سعود.
- بصيص، الطاهر. (٢٠٠٧). الخطاب الصحفي الجزائري إزاء انتفاضة الأقصى: دراسة تحليلية مقارنة بين الصحف الحكومية والصحف المستقلة من ٢٨/ ٩/ ٢٠٠٠ إلى ٣٠/ ٤/ ٢٠٠٣. رسالة ماجستير غير منشورة. جامعة القاهرة.
- البغدادي، ريما (٢٠١٠). المعالجة الإعلامية في قناتي الجزيرة والعربية للعدوان الإسرائيلي على غزة: دراسة مقارنة. رسالة ماجستير، الجامعة الخليجية في البحرين.
- خطابي، محمد (٢٠١٢). لسانيات النص: مدخل إلى انسجام الخطاب. الدار البيضاء: المركز الثقافي العربي.
- الخطيب، فاطمة. (٢٠١١). أطر المعالجة الإعلامية لسياسات الرئيس الأمريكي باراك أوباما: دراسة مقارنة بين قناتي الجزيرة والحرة. حوليات آداب عين شمس – مصر، مج ٣٩، الصفحات ٢٨٩ – ٣٢٦
- ريبول، آن؛ وموشلير، جاك (٢٠١٤). هل من الضروري مواصلة نهجنا في تحليل الخطابات؟ ترجمة: حافظ إسماعيلي علوي، ومحمد الملاخ. مجلة جامعة أم القرى لعلوم اللغات وآدابها، ع (١٣)، الصفحات ١٩٥-٢٣٦.
- شارودو، باتريك ومنغنو، دومينيك (٢٠٠٨). معجم تحليل الخطاب. ترجمة: عبدالقادر المهيري وحمادي صمود. تونس: المركز الوطني للترجمة. (٢٠٠٢).
- الشمري، عقيل والثبيتي، عبدالمحسن (٢٠١٥). لسانيات المدونات: نماذج وتطبيقات في لغة الصحافة العربية. ضمن: صالح العصيمي (مح.)، المدونات اللغوية العربية: بناؤها وطرائق الإفادة منها. الرياض: مركز الملك عبدالله الدولي لخدمة اللغة العربية.
- شومان، محمد. (٢٠١١). إشكاليات تحليل الخطاب في الدراسات الإعلامية العربية

الدراسات المصرية نموذجا. التربية، البحرين، س ٩، ع ٣١، الصفحات ٧٨ – ٩٥.

- عبدالغني، أمين سعيد. (٢٠١١). الخطاب التلفزيوني في معالجة القضايا العربية بقناتي الجزيرة الإنجليزية و BBC الدولية. دراسات الطفولة – مصر، مج ١٤، ع ٥١، الصفحات ٥٥ – ٧٤.

- العصيمي، صالح (محرر). (٢٠١٥). المدونات اللغوية العربية: بناؤها وطرائق الإفادة منها. الرياض: مركز الملك عبدالله الدولي لخدمة اللغة العربية.

- عكاشة، محمود (٢٠٠٥). لغة الخطاب السياسي: دراسة لغوية تطبيقية في ضوء نظرية الاتصال. القاهرة: دار النشر للجامعات.

- فاركلوف، نورمان (٢٠٠٩). تحليل الخطاب: التحليل النصي في البحث الاجتماعي. ترجمة: طلال وهبة. بيروت: المنظمة العربية للترجمة.

- مجدوب، عزالدين (٢٠١١). المقدمة العامة. ضمن: عزالدين مجدوب (محرر)، إطلالات على النظريات اللسانية والدلالية في النصف الثاني من القرن العشرين. تونس: المجمع التونسي للعلوم والآداب والفنون.

- محسن؛ ماهيناز. (٢٠٠٩). علاقة أساليب توظيف اللغة بأطر تقديم الأحداث داخل التقارير الإخبارية: دراسة تطبيقية على الحرب الإسرائيلية على غزة. المجلة المصرية لبحوث الإعلام، جامعة القاهرة، ع (٣٣).

- مكانري، توني وهاردي، أندرو. لغويات المدونة الحاسوبية: المنهج والنظرية والتطبيق. ترجمة: سلطان بن ناصر المجيول. دار جامعة الملك سعود، الرياض، ٢٠١٦.

ثانيًا: المراجع الأجنبية

- Baker، P. (2012). Acceptable bias? Using corpus linguistics methods with critical discourse analysis. Critical Discourse Studies، 9(3): 247-256.

- Baker، P.، Gabrielatos، C.، Khosravinik، M.، Krzyżanowski، M.، McEnery، T.، & Wodak، R. (2008). A useful methodological synergy: Combining critical discourse analysis and corpus linguistics to examine discourses of refugees and asylum seekers in the UK press. Discourse & Society، 19(3): 273-306.



- Beard، A. (2000). The language of politics. London: Routledge.
- Beaugrande، R. D. (1997). The story of discourse analysis. In T. van Dijk (ed.). Discourse as structure and process (Vol. 1). London: Sage publications, pp. 35-62.
- Bell، A. (1991). The language of news media. Oxford، UK: Blackwell.
- Burge، T. (1992). Philosophy of language and mind: 1950-1990. The Philosophical Review، 101(1): 3-51.
- Chilton، P. (2004). Analysing political discourse: Theory and practice. London: Routledge.
- Chouliaraki، L. (2008). Discourse analysis. In T. Bennett & J. Frow (eds.). The sage handbook of cultural analysis. London: SAGE Publications, pp. 674-698.
- Fairclough، N. L. (1985). Critical and descriptive goals in discourse analysis. Journal of pragmatics، 9(6): 739-763.
- Fairclough، N. (1989). Language and power. London: Longman.
- Fowler، R. (1991). Language in the news. Discourse and ideology in the British press. London: Routledge.
- Fowler، R. Hodge، R.، Kress، G.، & Trew، T. (1979). Language and control. London: Routledge & Kegan Paul.
- Gabrielatos، C.، & Baker، P. (2008) Fleeing، Sneaking، Flooding: A Corpus Analysis of Discursive Constructions of Refugees and Asylum Seekers in the UK Press، 1996-2005. Journal of English Linguistics 36(1): 5-38.
- Gee، J. P. (2014). An Introduction to Discourse Analysis: Theory and Method (4th ed.). London: Routledge.
- Howarth، D. (2000). Discourse: Concepts in the social sciences. Bukingham: Open University.
- Jaworski، A.، & Coupland، N. (2006). Introduction: Perspectives on discourse analysis. In A. Jaworski & N. Coupland (eds.). The discourse reader. Oxon: Routledge, pp. 1-38.



- Jørgensen، M. & Phillips، L. (2002). Discourse analysis as theory and method. London: Sage Publications.

- Kandil, M., 2009. The Israeli-Palestinian Conflict in American, Arab, and British Media: Corpus-Based Critical. (Unpublished doctoral thesis). Georgia State University.

- Kress، G. & Hodge، R. (1979). Language as ideology. London: Routledge & Kegan Paul.

- Mills، S. (1997). Discourse: The new critical idiom. London: Routledge.

- Orpin، D. (2005) Corpus Linguistics and Critical Discourse Analysis: Examining the Ideology of Sleaze، International Journal of Corpus Linguistics 10(1): 37–61.

- Paltridge، B. (2012). Discourse analysis: An introduction (2nd ed.). Bloomsbury Publishing.

- Schiffrin، D. (1994). Approaches to discourse. Oxford: Blackwell.

- Stubbs، M. (1997). Whorf›s children: Critical comments on critical discourse analysis (CDA). In A. Ryan and A. Wray (eds.). Evolving models of language. Clevedon: Multilingual Matters.

- van Dijk، T. (1988). News as discourse. Hillsdale، NJ: Erlbaum.

- van Dijk، T. (1993). Principles of critical discourse analysis. Discourse & Society، 4(2): 249-283.

- van Dijk، T. (1997). The study of discourse. In T. van Dijk (ed.). Discourse as structure and process (Vol. 1). London: Sage publications, pp. 1-31.

- van Dijk، T. (2001). Critical discourse analysis. In D. Schiffrin; D. Tannen & H. Hamilton (eds.). The handbook of discourse analysis. Malden: Blackwell Publishing, pp. 352-371.

- van Dijk، T. (2009) News، discourse and ideology. In K. Wahl-Jorgensen & T. Hanitzsch (ed.). The handbook of journalism studies. NY.: Routledge, pp. 191-204.



- van Leeuwen، T. (2009). Critical discourse analysis. In J. Renkema (ed.). Discourse، of Course: An overview of research in discourse studies. Amsterdam: Johns Benjamins. P. 277-292.

- Volosinov، V. (1973). Marxism and the Philosophy of Language. New York: Seminar Press.

- Weiss، G. & Wodak، R. (2003). Introduction: theory، interdisciplinarity and critical discourse analysis. In G. Weiss & R. Wodak (ed.). Critical discourse analysis: Theory and interdisciplinarity. NY.: Palgrave Macmillan, pp. 1-34.

- Widdowson، H. (1995). Discourse analysis: A critical view. Language and Literature، 4(3): 157-172.

- Wilson، J. (2001). Political discourse. In D. Schiffrin; D. Tannen & H. Hamilton (eds.). The handbook of discourse analysis. Malden: Blackwell Publishing, pp. 398-417.

- Wodak، R.، & Meyer، M. (2009). Critical discourse analysis: history، agenda، theory، and methodology. In R. Wodak & M. Meyer (eds.). Methods for critical discourse analysis. London: Sage Publications, pp. 1-33.


هذه الطبعة
إهداء من المركز
ولا يسمح بنشرها ورقياً
أو تداولها تجارياً

تعريف بالمشاركين في الكتاب

د. محمود بن عبدالله المحمود

أستاذ التخطيط والسياسة اللغوية المساعد

معهد اللغويات العربية-جامعة الملك سعود

الرياض (المملكة العربية السعودية)

mahm.almahmoud@gmail.com

أستاذ التخطيط اللغوي المساعد، ووكيل الشؤون الأكاديمية بمعهد اللغويات العربية - جامعة الملك سعود. حصل على درجة الدكتوراه في اللغويات (السياسة اللغوية والتخطيط اللغوي) من جامعة مكواري، بسدني-أستراليا. وحصل على درجتي ماجستير إحداها من أستراليا في اللغويات التطبيقية (إدارة البرامج اللغوية)، والأخرى من السعودية في اللغويات التطبيقية (تدريب معلمي العربية كلغة ثانية)، بالإضافة إلى دبلوم عال في تعليم العربية لغير الناطقين بها. كما يعمل مستشارا غير متفرغ لدى مركز الملك عبدالله لخدمة اللغة العربية في لجنة السياسة اللغوية والتخطيط اللغوي، ومستشارا في برنامج العربية للجميع، ويدير حاليا تحرير مجلة



التخطيط والسياسة اللغوية. قدم العديد من أوراق العمل في مؤتمرات تخصصية، وقدم العديد من ورش العمل لمعلمي العربية لغير الناطقين بها في دول مختلفة. وتتضمن اهتماماته البحثية اللغويات الاجتماعية، والسياسة اللغوية والتخطيط اللغوي، وإدارة البرامج اللغوية، وتدريب معلمي اللغة الثانية وإعدادهم، وتعليم العربية لغةً ثانية. (انظر: http://fac.ksu.edu.sa/mmahmoud/cv).

د. عقيل بن حامد الشمري

أستاذ اللغويات التطبيقية المساعد

معهد اللغويات العربية-جامعة الملك سعود

الرياض (المملكة العربية السعودية)

alzammai3@yahoo.com

أستاذ اللسانيات التطبيقية المساعد، ورئيس قسم تدريب المعلمين بمعهد اللغويات العربية، بجامعة الملك سعود، الرياض. حصل على درجتي الماجستير والدكتوراه في جامعة كوينزلاند، أستراليا. اهتماماته البحثية تشتمل على اكتساب اللغة والحرافة literacy، وأبحاث الكتابة باللغة العربية، وتعلم اللغة بمساعدة الحاسوب. عمل محاضرا في تدريس مادة «اكتساب اللغة الثانية» لطلاب الماجستير في جامعة كوينزلاند، بالإضافة إلى تدريس مادة «الحاسوب واللغة» لطلاب الماجستير في معهد اللغويات العربية. يعمل في مجال تعليم اللغة العربية لغير الناطقين بها، وشارك في العديد من الدورات التدريبية لمعلمي العربية في عدد من الدول المختلفة. عضو في لجنة التخطيط والسياسة اللغوية بمركز الملك عبدالله الدولي لخدمة اللغة العربية. وعمل أيضا مستشارا غير متفرغ في مشروع الملك عبدالله بن عبدالعزيز لتطوير التعليم العام. (انظر: http://fac.ksu.edu.sa/alaqeel/cv).

د. هند بنت مطلق العتيبي

أستاذ اللغويات التطبيقية والترجمة المشارك

كلية اللغات والترجمة
جامعة الملك سعود

الرياض، المملكة العربية السعودية

hialotaibi@ksu.edu.sa

وكيلة كلية اللغات والترجمة سابقاً، أكاديمية وباحثة حاصلة على شهادة البكالوريوس في تطبيقات الحاسب الآلي من جامعة الملك سعود، والماجستير والدكتوراه في تعليم اللغات بواسطة الحاسب من جامعة مانشستر في بريطانيا. تعمل حالياً كأستاذ مساعد في كلية اللغات والترجمة جامعة الملك سعود. تشمل اهتماماتها البحثية مجالات تقنيات التعليم، التعليم الإلكتروني والتعليم عن بعد، وتطبيقات الحاسب في الترجمة وتعليم اللغات. (انظر: http://fac.ksu.edu.sa/hialotaibi).

د. عبدالله بن يحيى الفيفي

أستاذ اللغويات الحاسوبية المساعد

معهد تعليم اللغة العربية
جامعة الإمام محمد بن سعود الإسلامية

الرياض، المملكة العربية السعودية

ayjfaifi@imamu.edu.sa

أستاذ الدراسات العليا المساعد للغويات الحاسوبية في جامعة الإمام محمد بن سعود الإسلامية، الرياض. درس البكالوريوس في اللغة العربية في قسم اللغة والأدب في جامعة الملك خالد، ودرس الماجستير في تعليم اللغة بمساعدة الحاسب في قسم اللغويات في جامعة Essex، والدكتوراه في اللغويات الحاسوبية ومعالجة اللغة الطبيعية في قسم الحاسب الآلي في جامعة Leeds، وكلاهما في المملكة المتحدة. شارك



في عدد من المشاريع العلمية في مجال معالجة اللغة العربية حاسوبياً من أبرزها: المدونة اللغوية لمتعلمي اللغة العربية، ومعجم «شائع» لمتعلمي اللغة العربية الناطقين بغيرها في المستوى المبتدئ. له عدة أبحاث منشورة حول حوسبة اللغة العربية، والمدونات اللغوية، والمعاجم الإلكترونية، والتعليم عن بعد، كما أسهم في تأليف عدد من الكتب المنشورة في هذه المجالات. يعمل مستشاراً ومحكماً لمجموعة من الجهات والدوريات العلمية والمؤتمرات الدولية. مهتم بعلم اللغة الحاسوبي، وتعليم اللغة بمساعدة الحاسب.(انظر: (https://members.imamu.edu.sa/staff/ayjfaifi/Pages/default.aspx

د. عبدالمحسن بن عبيد الثبيتي

أستاذ الحوسبة المشارك
المركز الوطني لتقنية الحاسب والرياضيات التطبيقية
مدينة الملك عبدالعزيز للعلوم والتقنية
الرياض، المملكة العربية السعودية
aalthubaity@kacst.edu.sa

أستاذ البحث المشارك في المركز الوطني لتقنية الحاسب والرياضات التطبيقية في مدينة الملك عبدالعزيز للعلوم والتقنية، وباحث رئيس لمشروع المدونة اللغوية العربية لمدينة الملك عبدالعزيز للعلوم والتقنية King Abdulaziz City for Science and Technology's Arabic Corpus KACSTAC. عمل في مجال إدارة تقنية المعلومات لعشرين عامًا قبل انتقاله للعمل في مدينة الملك عبدالعزيز للعلوم والتقنية. حصل على الدكتوراه من جامعة ساري ببريطانيا من قسم الحوسبة، ونشر العديد من الأبحاث في المؤتمرات والدوريات المحكمة. مهتم بالمصادر اللغوية العربية ومعالجة اللغة العربية والتنقيب في النصوص والتعقب الآلي لتطور المعرفة في المجالات المتخصصة. (انظر: .(https://www.researchgate.net/profile/Abdulmohsen_Al-thubaity

د. محمود عثمان الحاج

أستاذ بحث الحوسبة مشارك
كلية الحوسبة والاتصالات
جامعة لانكستر
المملكة المتحدة
m.el-haj@lancaster.ac.uk

باحث مشارك أول في كلية الحاسبات والاتصالات في جامعة لانكستر. حصل على درجة الدكتوراه في علوم الحاسب الآلي من جامعة إسيكس في بريطانيا. باحث في مجال التلخيص الآلي للغة العربية. تشمل اهتماماته البحثية مجالات عدة منها معالجة اللغات الطبيعية واللغة العربية، والتعلم الآلي، واستخراج المعلومات، واللسانيات الحاسوبية للغة الإنجليزية والعربية في مجال المحاسبة والإدارة والتقارير السنوية. وقد عمل على عدة مشاريع بحثية متعددة التخصصات في جامعة لانكستر وعن طريق التعاون مع كبرى الشركات المالية والمصرفية في لندن، وقد عمل سابقا مطورَ برامج وباحث في مجال التنقيب في البيانات فترة عمله في مؤسسة أرشيف بيانات المملكة المتحدة. (انظر: http://www.lancaster.ac.uk/staff/elhaj/).

د. سلطان بن ناصر المجيول

أستاذ لغويات المدونة الحاسوبية المساعد
قسم اللغة العربية-كلية الآداب
جامعة الملك سعود
الرياض، المملكة العربية السعودية
salmujaiwel@ksu.edu.sa

أستاذ مساعد في لغويات المدونة الحاسوبية والتحليل اللغوي الآلي والإحصائي. حصل على شهادة الدكتوراه من جامعة إكسيتر في الدراسات المتداخلة اختصاصيًا



interdisciplinary studies حول تحليل المداخل المعجمية الإنجليزية/ العربية الغامضة والخاطئة آليا. يعمل محررًا لمجلة الدراسات العربية الصادرة عن دار نشر: تايلور وفرانسيس بأبينجدون Abingdon ببريطانيا والتي تُنشر باللغة الإنجليزية http://www.tandfonline.com/loi/rjab20. قام بتحكيم كثير من الأبحاث في مجال التحليل التركيبي ولغويات المدونة الحاسوبية. له العديد من المشاركات في المؤتمرات المتعلقة بلغويات المدونة الحاسوبية واللغويات التطبيقية. مهتمٌ بالدراسات ذات التداخل الاختصاصي، وبخاصة المجال التحليلي الإحصائي للغة. تتنوع أبحاثه ذات التداخل الاختصاصي بالمناهج اللغوية المدونية الحاسوبية لغرض تحليل النصوص اللغوية (اللسانية) في التاريخ التحليلي historiography والاتجاهات الحديثة new trends، والعلمية، والأكاديمية، والاقتصادية، والاجتماعية (انظر: http://fac.ksu. edu.sa/salmujaiwel/publications).


هذه الطبعة
إهداء من المركز
ولا يسمح بنشرها ورقياً
أو تداولها تجارياً

الفهرس


هذه الطبعة
إهداء من المركز
ولايسمح بنشرها ورقياً
أو تداولها تجاريا


هذه الطبعة
إهداء من المركز
ولا يسمح بنشرها ورقياً
أو تداولها تجاريا


هذه الطبعة
إهداء من المركز
ولايسمح بنشرها ورقياً
أو تداولها تجارياً


هذه الطبعة
إهداء من المركز
ولا يسمح بنشرها ورقياً
أو تداولها تجاريا


هذه الطبعة
إهداء من المركز
ولايسمح بنشرها ورقياً
أو تداولها تجاريا



مباحث لغوية ٣٦

لغويات المدونة الحاسوبية

تطبيقات تحليلية على العربية الطبيعية

اعتنى هذا الكتاب بمعالجة النصوص العربية آليًا. واحتوى على ستة فصول: كل فصل قد وظّف مدونةً حاسوبية معينة بأدوات منهجية لغوية حاسوبية متعددة.

وتنوعت الفصول في المناهج والأدوات بالصورة التي تجعل من كل فصل مُمَثِّلاً لمنهج ونظرية وتطبيق من عدة مناهج ونظريات وتطبيقات ضمن الحقل نفسه.

وقد تطلّب التأليف في هذا العلم الكثير من الجهد والوقت والتحليل العلمي من المشاركين، ووفقَ مناهج (لغويات المدونة الحاسوبية)، ولذلك فإن طلبَ الاستفادة منه يتطلب أيضا وقتًا يمتد كامتداد الجهد المبذول فيه؛ لأنه يحتاج إلى تمرّس دائم على مناهج الإحصاء المتعددة والمعالجة الحاسوبية اللغوية ضمن نظريات لغويات المدونة الحاسوبية، ومناهجها، وتطبيقاتها.

وليس الغرض من الكتاب حصر المعالجات الآلية، كونها واسعةَ المجال ومتشعبة الموضوعات بتشعّب أغراض المعالجة، إنما الغرض التعريف بطبيعة التطبيقات.

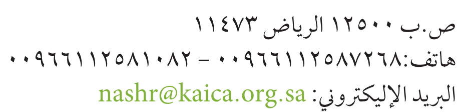
والمجال مفتوح للقراء والمختصين في هذا المجال بطلب بيانات كل تحليل؛ من أجل العمل على مبادئ التعليلية accountability أو الدحضية falsifiability أو التناتجية (التماثل والقبول) replicability العلميّة دون غيرها، والتواصل مع المشاركين عبر حساباتهم البريدية المشار إليها في (التعريف بالمشاركين)؛ من أجل الحصول على بيانات التحليل.

محرر الكتاب

د. سلطان بن ناصر المجيول






ص.ب ١٢٥٠٠ الرياض ١١٤٧٣
هاتف:٠٠٩٦٦١١٢٥٨٧٢٦٨ - ٠٠٩٦٦١١٢٥٨١٠٨٢
البريد الإليكتروني: nashr@kaica.org.sa


9 786039 084792